

Identification of Specific Language Impairment in
simultaneous and (early) successive bilingual
children in Germany:
Studies on typical and atypical bilingual language
development

Thesis approved by the School of English and American Studies
-Faculty III-
of the University of Oldenburg
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy (Dr. Phil.)

by Ms. **Lina Abed Ibrahim**
born on 3. June 1984 in Charkiw, Ukraine

Reviewer Prof. Dr. Cornelia Hamann

Co-reviewers Prof. Dr. Angela Grimm

Prof. Dr. Juliane Goschler

Date of thesis defense 14.12.2023

To mama and 𐤁𐤏𐤃, for the gift of multilingualism.

Abstract

Although bilingualism neither causes nor predisposes children to language disorders, bilingual children are frequently misdiagnosed with specific language impairment (SLI) due to the paucity of appropriate assessment tools. This dissertation aims to inform evidence-based language assessment of multilingual children in Germany by evaluating the clinical potential of heritage language (HL) assessment and of newly developed LITMUS sentence and nonword repetition tasks for identifying SLI in preschool and primary-school-age (ages: 5;6-8;11 years) monolingual, simultaneous, and early successive bilingual children acquiring German as a societal language. In light of the findings of the studies presented in this thesis, it can be concluded that standardized HL assessment, like standardized assessment in the societal language, poses a high risk of overidentifying heritage typically developing bilingual children as language-impaired, even when pathology cut-offs are adjusted for language dominance. By contrast, owing to their careful construction and cross-linguistic availability, LITMUS sentence- and nonword-repetition tasks targeting aspects of syntactic and phonological complexity provide time-efficient, easy-to-administer measures that reliably distinguish typical from atypical bilingual language development in heritage bilingual children, not only in the societal language but also in the heritage/minority language, especially when tests targeting different language components are combined.

Keywords: Specific Language Impairment, Developmental Language Disorder, bilingual language acquisition, sentence repetition task, nonword repetition task

Funding information:

This research was carried out between 2014 and 2020 at the University of Oldenburg within the collaborative projects BiLaD (Bilingual Language Development: Typically Developing Children and Children with Language Impairment) and BiliSAT (Bilingual Language Development in School-age Children with/without Language Impairment with Arabic and Turkish as first languages). The BiLaD project was financed by a joint grant from the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation; HA 2335/6-1, CH 1112/2-1, and RO 923/3-1) and a French ANR grant (ANR-12-FRAL-0014-01). The BiliSAT project was funded by DFG Grants to CH 1112/4-1S (to Chilla) and HA 2335/7-1C (to Hamann).

Acknowledgments

The completion of this dissertation marks the end of an intense and deeply rewarding chapter in my academic, professional, and personal life. This journey would not have been possible without the support, guidance, and encouragement of many individuals, to whom I extend my heartfelt thanks. In the next few paragraphs, I would like to acknowledge some of the remarkable people who have made this dissertation a reality.

First and foremost, I would like to express my deepest gratitude to my supervisor and one-of-a-kind linguist, Cornelia Hamann, who, thirteen years ago, introduced me to a field of research I had never imagined pursuing, inspired my work, consistently helped me keep perspective, and taught me far more than she will ever know. *Cornelia*, it was the greatest honor to be your last doctoral student. I can hardly express my gratitude for your extraordinary mentorship and for being a dear friend all these years. I cannot thank you enough for your unconditional support and for the countless hours you patiently spent discussing my research, even on vacation in Sweden or in the mountains, and even after your retirement. You were a *Doktormutter* to me in the literal sense of this word. One day, I hope to be as great a mentor to my students as you have been to me.

I am also grateful to the other members of the thesis examination committee, Angela Grimm, Juliane Goschler, Esther Ruigendijk, and Ron Geluykens, and of the doctoral colloquium of Faculty III at the University of Oldenburg. Your incisive comments, constructive feedback, and thoughtful suggestions, now integral to this thesis, have made it much stronger than it would have been otherwise. I am particularly indebted to István Fekete, one of the best statistical experts I know, for his support with complex statistical analyses, especially the machine-learning algorithms employed in Study 2. The latter study would not have existed without his support. A very special thank-you goes to Ilka Flöck, Dobrinka Genevska-Hanke, Irina Hertel, and the members of the doctoral colloquium led by Solveig Chilla at the Institute for Special Education, Europa-Universität Flensburg, for their emotional support and insightful feedback during my preparations for the thesis defense. I am also grateful to the DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) for funding the BiLaD and BiliSAT projects, which supported the PhD positions I held during crucial stages of my dissertation, and the Kind-Hörstiftung, whose prize helped make the printing of this dissertation possible.

I owe a huge debt to Solveig Chilla, whom I have had the pleasure of working for not only as a co-lead on the DFG-funded projects BiLad and BiliSAT, within which this thesis arose, but also as my supervisor at the Department of Speech and Language Education, Europa-Universität Flensburg, for the past five years. Thank you so much for enabling a seamless transition in my academic career and for securing for me the time, financial means, and space to complete this dissertation; without your support, writing up and revising it would not have been possible.

An empirical thesis would not exist without empirical data; therefore, I would like to thank the participating children, their parents, and the directors of the participating schools and speech and language therapy centers, especially Martina Natour, Ulla Wortman, and the AWO institutions in Bremen and Lower Saxony, for their steadfast support with participant recruitment. Special thanks go to the BiLaD and BiliSAT projects' former student and research assistants, especially Marie Schnieders and David Öwerdieck, for assisting with data collection and processing. I will never forget the late-night shifts you spent working on the data to help us meet an abstract submission deadline. I would also like to thank the former BiLaD and BiliSAT team members in Germany and France: Monika Rothweiler, Tatjana Lein, Hilal Şan, Juliane Maier, Laurice Tuller, Philippe Prévost, Racha Zebib, Sandrine Ferré, Christophe dos Santos, and Laetitia de Almeida. I benefited greatly from our fruitful discussions and joint publications, and I learned a great deal from each of you.

Last but not least, my deepest thanks to the most wonderful team I will ever be part of: my dearest family. To my father Ayoub, who, as far as I can remember, taught me the value of hard work and of seeking knowledge, even if that meant literally crossing mountains. I will never forget how you risked your life during the war to bring me across the border so I would not miss the fall semester. To my mother Tatjana, for being my safe harbor and for cheering me up during stormy times, and those were not few. To my second mother, Leila, my aunt Olga, and my little brother, Aladdin (who is not little anymore), for the countless hours they spent babysitting my children so that I could finalize the dissertation. To the best partner and husband I could have hoped for, Walid Al Otaibi: being on a team with you turned moments that would have been intense challenges into a memorable journey. Thank you for always believing in me and standing by me through every high and low, even when work kept me up late into the night, and deadlines meant I had to miss family time. I can never make this up to you. Thank you to my beautiful children, Khalid and Bella Zeyne, for all the joy you brought to my life, and for driving

me to be a better version of myself every day. Khalid, thank you for commuting with me to all possible national and international conferences and for enduring countless late afternoons in daycare, just because Mom had another publication deadline!

Finally, my thanks would be incomplete without honoring the one who taught me the art of resilience and thriving under the most difficult circumstances: my beloved *Palestine*.

German Summary

Zusammenfassung

Aktuelle Spracherwerbsforschung hat sich auf die Problematik bei der Diagnose einer Spezifischen Sprachentwicklungsstörung (SSES), bzw. einer Sprachentwicklungsstörung (SES) im Kontext von Mehrsprachigkeit fokussiert (siehe Bishop et al., 2017 und Rice, 2020a für eine Diskussion). Eine der Hauptursachen für die Schwierigkeiten der Diagnostik einer SES bei monolingualen Kindern ist die erhebliche Heterogenität innerhalb der sprachlichen Profile, wobei die Schwere der Störung innerhalb der verschiedenen Komponenten variieren kann (Conti-Ramsden et al., 2001a; Crutchley et al., 1997; Friedmann & Novogrodsky, 2011; Leonard, 2014a). Neben den Herausforderungen in der Diagnostik, die durch die Heterogenität der Störung selbst entstehen, gewinnt die Identifikation einer Sprachentwicklungsstörung bei bilingualen Kindern zusätzlich an Komplexität durch die große Variabilität innerhalb der (typischen) Sprachentwicklung, geprägt durch eine Vielzahl Sprachinterner und -externer Faktoren (Birdsong, 2018; Grüter & Paradis, 2014; Hamann, 2012; Paradis, 2023). Die Unterscheidung zwischen einer tatsächlich vorliegenden Sprachentwicklungsstörung und mangelnden Sprachkenntnissen aufgrund eines unzureichenden Sprachkontakts, wird auch dadurch erschwert, dass sich die sprachlichen Profile von ein- und mehrsprachigen Kindern mit und ohne Sprachentwicklungsstörung teilweise überschneiden.

Der Anteil von Kindern, die in Deutschland mehrsprachig aufwachsen, steigt stetig an. Für viele Kinder findet jedoch der erste systematische Kontakt mit der Mehrheitssprache erst in der (Vor)Schule statt. Insbesondere im Fall einer sequentiellen bilingualen Entwicklung befinden sich die Kinder zu dem Zeitpunkt, an dem ihr Sprachstand ermittelt wird, noch in einer Phase, in der sich die Sprachfähigkeiten in der Mehrheitssprache entwickeln (Paradis, 2016). Dies hat zur Folge, dass sie in standardisierten Sprachstanderhebungen signifikant schlechtere Ergebnisse erzielen als einsprachige Kinder. Dies führt sowohl zu Überdiagnosen als auch zu Unterdiagnosen von Spracherwerbsstörungen bei bilingualen Kindern (Grimm & Schulz, 2014, 2017), mit nachteiligen Folgen für die Kinder selbst, ihre Familien und die Gesellschaft insgesamt. Einer der Gründe für Fehldiagnosen ist der Mangel an standardisierten diagnostischen Verfahren für bilinguale Kinder, der auf die zurzeit noch unvollständigen Kenntnisse über den typischen Spracherwerb bei mehrsprachigen Kindern zurückzuführen ist.

Jedoch ist eine korrekte Sprachstanderhebung essentiell für weitere Entscheidungen darüber, ob dem Kind Zugang zu logopädischen Angeboten gewährt werden kann, die über die Sprachförderung an Bildungseinrichtungen hinaus gehen (Paradis et al., 2008). Entscheidungen dieser Art sind umso relevanter, da Kinder mit Sprachentwicklungsstörungen auf Behandlungen angewiesen sind, die sich von anderweitig benachteiligten Kindern mit intakten Sprachfähigkeiten unterscheiden (Roy et al., 2014). Zudem spielt die korrekte Ermittlung des Sprachstandes eine integrale Rolle für die weitere Therapie und für das Zuschneiden von Behandlungen, die den Bedürfnissen des Kindes entsprechen (De Lamo White & Jin, 2011).

Als mögliche Lösung für das Vermeiden von Fehldiagnosen wurde vorgeschlagen, das Kind mindestens in der dominanten Sprache (Fredman, 2006), besser jedoch in beiden Sprachen zu bewerten (ASHA, 2023; IALP, 2011; RCSLT, 2007), da sich eine echte Sprachentwicklungsstörung in beiden Sprachen manifestiert. Im klinischen Alltag ist ein solches Vorgehen jedoch oft nicht realisierbar und evidenzbasierte Empfehlungen für die Interpretation der Testergebnisse mehrsprachiger Kinder sind in unzureichender Anzahl vorhanden (Peña et al., 2016). Eine weitere Möglichkeit sieht vor, diagnostische Instrumente für einsprachige Populationen gemäß den Fähigkeiten ihrer jeweils dominanten Sprache anzupassen, siehe Empfehlungen von Thordardottir (2015a). Allerdings gilt dieser Vorschlag ausschließlich für simultan bilinguale Kinder und eignet sich womöglich nicht für Kinder mit anderen Erwerbsszenarien.

Während der COST Action IS0804 “Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment” wurden mehrere sprachübergreifende LITMUS (**L**anguage **I**mpairment **T**esting in **M**ultilingual **S**ettings) Verfahren entwickelt, um den diagnostischen Herausforderungen in mehrsprachigen Kontexten gerecht zu werden (siehe Armon-Lotem et al., 2015a). Von besonderer Relevanz für die vorliegende Arbeit sind der Satzwiederholungstest (LITMUS-SRT, Hamann et al., 2013) und der quasi-universelle Kunstwortwiederholungstest für Deutsch (LITMUS-QU-NWRT, Grimm et al., 2014). Solche Wiederholungstests finden regelmäßig und erfolgreich Anwendung in der Diagnostik von SES in monolingualen Populationen (Conti-Ramsden et al., 2001a). Während Satzwiederholung morphosyntaktische Kompetenzen erfasst (Chiat et al., 2013; Fleckstein et al., 2018; Hamann et al., 2017; Polišenská et al., 2015; Vinther, 2002), liegt der Fokus beim Kunstwortwiederholen auf phonologischen Strukturen (Chiat, 2015; Chiat & Polišenská,

2016; dos Santos & Ferré, 2018; Grimm, 2022), die Kinder mit SES vor besonders hohe Herausforderungen stellen (Leonard 2014a). Das Entscheidende bei diesen Verfahren ist, dass sie Strukturen einbinden, die besonders Kindern mit Sprachentwicklungsstörung unabhängig ihrer jeweiligen Herkunftssprache Schwierigkeiten bereiten, nicht jedoch Kindern, die zwei- oder mehrsprachig aufwachsen.

Neuste Studien mit LITMUS-Verfahren liefern vielversprechende Ergebnisse in mehrsprachigen Kontexten. So unterscheiden die beiden Wiederholungstests erfolgreich, und in Kombination besonders verlässlich, mono- und bilinguale Kinder mit und ohne SES voneinander (u.a. Armon-Lotem & Meir, 2016; Chilla et al., 2021; de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018; siehe auch Grohmann & Armon-Lotem, 2021). Keine dieser Studien hat jedoch untersucht, ob sprachliche Phänomene des bilingualen Erwerbs Einfluss auf die Testergebnisse in einem Ausmaß haben könnten, welches deren diagnostische Genauigkeit erheblich beeinträchtigen könnte. Darüber hinaus gibt es wenig Studien, die den Einsatz der LITMUS Verfahren in der Minderheitssprache (Herkunftssprache) erprobt hatten.

Die vorliegende Dissertation ist der Schnittstelle zwischen Mehrsprachigkeit und Sprachentwicklungsstörung zuzuordnen und entstand vorwiegend im Rahmen des deutsch-französischen Kooperationsprojekts BiLaD (Bilingual Language Development: Typically Developing Children and Children with Language Impairment). Hauptanliegen des Projekts war die Entwicklung sprachübergreifender Kriterien für die Diagnostik von SES bei mehrsprachigen Kindern abseits ihrer Fähigkeiten in der Erst- und Zweitsprache, während das konkrete Ziel darin bestand, die entwickelten LITMUS-Verfahren in ihren soziolinguistisch differierenden deutsch-französischen Kontexten zu validieren.

Die Dissertation besteht aus drei miteinander verzahnten Studien. Die erste Studie setzt sich mit der Empfehlung auseinander, beide Sprachen eines bilingualen Kindes zu testen. Am Beispiel des Arabischen als Herkunftssprache in Deutschland wird untersucht, ob eine standardisierte Testbatterie in der Erstsprache und ein LITMUS-Satzwiederholungstest für Arabisch für die erfolgreiche Diagnostik von SES in simultan und früh-sukzessiven bilingualen Kindern mit Arabisch als Erstsprache herangezogen werden können. Es wird zudem der Frage nachgegangen, welche Faktoren, wie z.B. Alter und Kontaktdauer, die Varianz der Leistungen in den verschiedenen sprachlichen Bereichen erklären. Die Ergebnisse zeigten, dass trotz Anpassung von monolingualen Normierungen für bilinguale Populationen ein wesentlicher Teil der simultan/früh-sukzessiven Gruppe im standardisierten Testverfahren mindestens zwei

unterdurchschnittliche Resultate im pathologischen Bereich erzielten. Besonders benachteiligt waren die Kinder im rezeptiven und expressiven Wortschatz sowie bei der Produktion morphosyntaktischer Strukturen, mit negativen Effekten von Kontaktdauer und L2-Sprachgebrauch. Das bedeutet, dass alleiniges Vertrauen auf standardisierte L1 Verfahren für die Feststellung einer Sprachentwicklungsstörung anhand der Herkunftssprache nicht möglich ist. Bei einem L1 LITMUS-Satzwiederholungstest zeigte sich ein anderes Bild: Im Gegensatz zu Kindern mit SES erzielte die simultan/früh-sukzessiv bilinguale Gruppe mit typischer Sprachentwicklung ähnliche Resultate wie die monolingualen und spät-sukzessiv bilingualen Kinder ohne Sprachstörung im arabischsprachigen LITMUS-Satzwiederholungstest, was darauf hinweist, dass der Test für simultane und früh-sukzessive mehrsprachige Populationen durchaus angemessen ist.

Die zweite Studie fokussiert die Sprachdominanz und befasst sich mit der Frage, ob diese einen Einfluss auf die diagnostische Genauigkeit der deutschen Varianten des LITMUS-SRTs und QU-NWRTs hat. Ein unüberwachter maschineller Lernalgorithmus kam zum Einsatz, um eine klinische Kategorie für die Kinder als $\pm$ sprachgestört auf Grundlage ihrer Ergebnisse beim SRT und QU-NWRT (einzeln und in Kombination) abzuleiten. Informationen über den klinischen Status auf Grundlage einer standardisierten Beurteilung in beiden Sprachen wurde währenddessen zurückgehalten. Anschließend wurde die diagnostische Genauigkeit berechnet und es wurde mit Hilfe einer Regressionsanalyse untersucht, welche Hintergrundvariablen die Zugehörigkeit zu der klinischen Gruppe am besten erklärten. Die Ergebnisse zeigen, dass Risikofaktoren für SES in allen Modellen signifikante Prädiktoren für die Zugehörigkeit zur klinischen Gruppe waren. Sprachdominanz hingegen trug nicht zur Erklärung der Varianz bei. Dies bestätigt, dass die diagnostische Genauigkeit der Testverfahren nicht durch die Sprachdominanz beeinträchtigt wird, auch wenn sie einen offenkundigen Einfluss auf die Performanz von typisch entwickelten bilingualen Kindern im SRT hat. Die Ergebnisse zeigen außerdem, dass eine Kombination von SRT, bewertet nach Zielstruktur, und dem phonologisch komplexeren sprachabhängigen Teil des QU-NWRTs, höhere diagnostische Genauigkeit ergibt als jeweilige Einzelmaßnahmen.

Da syntaktische Komplexität eine zentrale Rolle bei der Konzeption und Konstruktion der deutschen Version des LITMUS-SRTs spielte, wurde abschließend das Zusammenspiel zwischen syntaktischer Komplexität und Wiederholungsgenauigkeit in unterschiedlichen Gruppen (mono- und bilingual, mit und ohne SES) einer größer angelegten Stichprobe untersucht. Von Interesse war auch, ob die Varianz der Ergebnisse

primär durch sprachliche Fähigkeiten (lexikalische und morphosyntaktische Kompetenzen, Klem et al., 2015) oder durch das Arbeitsgedächtnis (verbales Kurzzeitgedächtnis und verbales Arbeitsgedächtnis) erklärt werden kann. Angesichts der Tatsache, dass Kinder mit SES Defizite in linguistischen Repräsentationen und deren Ableitungsoperationen/Derivationen haben, die mit einer begrenzten Kapazität des Arbeitsgedächtnisses interagieren, ging die Studie davon aus, dass sie große Schwierigkeiten mit Konstruktionen, die Aspekte syntaktischer Komplexität beinhalten, wie z.B. (*wh*-)Bewegung, syntaktische Einbettung und Intervention (Friedmann et al., 2009; Hamann et al., 2017). Die Ergebnisse zeigen gute diagnostische Genauigkeit, sowohl für monolinguale als auch für bilinguale Kinder. Die Ergebnisse zeigen ferner, dass die Wiederholungsgenauigkeit mit zunehmender syntaktischer Komplexität sowohl bei typisch entwickelten als auch bei sprachgestörten-Kindern abnimmt. Der Effekt der Komplexität in den monolingualen und bilingualen Gruppen mit SES ist hingegen deutlich ausgeprägter, zumal die Kinder auf allen Ebenen signifikant schlechtere Leistungen erbringen als die monolingualen und bilingualen Kinder mit typischer Sprachentwicklung und bei Strukturen mit mehr als einem Komplexitätsfaktor sogar nahezu Nullperformanz zeigen. Dies bestätigt, dass der LITMUS-SRT nicht nur das Arbeitsgedächtnis, sondern die wesentlichen morphosyntaktischen Defizite von SES erfasst. Dennoch erwiesen sich einige Strukturen, wie z.B. Objekt-Topikalisierungen und Objekt-Relativsätze mit Intervention als besonders schwierig auch für die typisch entwickelten Kinder in der untersuchten Altersgruppe, so dass sie nicht zur Identifikation besonderer sprachlicher Probleme dienen können. Dies deutet darauf hin, dass solche Strukturen gemieden werden sollten, vor allem in SRTs, die bei jüngeren bilingualen Kindern eingesetzt werden sollen.

Die in dieser Dissertation vorgestellten Studien liefern eine gründliche Evaluierung des klinischen Potenzials der standardisierten L1-Bewertung sowie der LITMUS-Satz- und Kunstwortwiederholungstests zur Identifizierung von SES bei bilingualen Kindern, die Deutsch als Gesellschaftssprache erwerben. Eine wichtige Implikation für die klinische Praxis ist, dass der Einsatz standardisierter Sprachverfahren in der Herkunftssprache ein hohes Risiko für Fehldiagnosen birgt; dies deutet darauf hin, dass direkte Sprachmessungen in der Mehrheitssprache eine bessere diagnostische Alternative darstellen, insbesondere bei simultan und fröhsukzessiv bilingualen Kindern. Die LITMUS-Wiederholungstests bieten dafür zeiteffiziente und leicht durchzuföhrnde Verfahren, um typischen vom atypischen Spracherwerb bei bilingualen Kindern zu

unterscheiden, besonders wenn Kombinationen von Verfahren eingesetzt werden, die jeweils verschiedene Sprachkomponenten erfassen.

Table of Contents

Abstract	II
Acknowledgments.....	III
German Summary	VI
Table of Contents	XII
List of Tables.....	XVI
List of Figures.....	XVIII
1. General introduction.....	1
1.1 Outline of this dissertation	7
2. Theoretical background	9
2.1 Bilingual child language acquisition.....	9
2.1.1 Sources of individual variation in bilingual language development.....	10
2.1.1.1 Child-internal factors	11
2.1.1.2 Child-external factors.....	20
2.1.1.3 Cross-linguistic influence (CLI)	24
2.1.2 Bilingual language development across language domains	36
2.1.2.1 Lexical abilities in typically developing bilingual children.....	37
2.1.2.2 Development of complex (morpho)syntax in bilingual children	40
2.1.2.3 Phonological abilities in bilingual children.....	60
2.1.3 Bilingualism and executive function	65
2.2 Specific Language Impairment (SLI)	77
2.2.1 Vulnerable domains in SLI.....	80
2.2.1.1 Morphosyntactic abilities of children with SLI	81
2.2.1.2 Phonological abilities of children with SLI	109
2.2.1.3 Lexical-semantic abilities in children with SLI	113
2.2.1.4 Executive functioning in children with SLI.....	120
2.2.2 Theoretical accounts for SLI	126
2.2.3 Long-term outcomes of SLI.....	130
2.3 Diagnostic challenges in assessing bilingual children	132
2.3.1 LITMUS tools for bilingual language assessment.....	135
2.3.1.1 Sentence repetition tasks	137
2.3.1.2 Nonword repetition tasks	142
2.3.1.3 The parental questionnaire (PaBiQ).....	148
2.3.2 Estimating diagnostic accuracy of screening tools	149
3. Materials and Methods	152
3.1 Participants.....	153

3.1.1	Verification of clinical group membership.....	154
3.2	Materials	157
3.2.1	Standardized L1 and L2 tests.....	157
3.2.1.1	Standardized L1 tests	159
3.2.1.2	Standardized tests for assessment of L2-German	164
3.2.2	Standardized tests for assessment of non-verbal intelligence and working memory.....	167
3.2.3	The LITMUS tools	168
3.2.3.1	The German LITMUS Sentence Repetition Task.....	168
3.2.3.2	The German LITMUS-QU-Nonword Repetition Task (LITMUS-QU-NWRT).....	174
3.2.3.3	The Questionnaire for parents of bilingual children LITMUS-PaBiQ	177
3.3	General procedure and statistical treatment	181
4.	Study 1: Assessment of Heritage Language abilities in child bilinguals with and without SLI: Comparing a standardized L1 test battery to an Arabic LITMUS-Sentence repetition task.....	182
4.1	Introduction.....	183
4.2	Child heritage language development.....	186
4.2.1	Sources of individual variation in heritage language development	187
4.2.1.1	Child-internal factors	187
4.2.1.2	Language environment factors	189
4.2.2	HL outcomes across linguistic domains	192
4.3	Arabic as a heritage language	196
4.3.1	Some notes on Arabic	196
4.3.2	Previous findings on the acquisition of Arabic as a heritage language	197
4.4	The present study	201
4.4.1	Research questions and aims of the present study	204
4.5	Method	205
4.5.1	Participants	205
4.5.2	Materials	209
4.5.3	Data analysis	217
4.6	Results.....	218
4.6.1	Standardized test for Arabic (ELO-L)	218
4.6.1.1	Lexical abilities	219
4.6.1.2	Morphosyntactic abilities.....	221
4.6.1.3	Phonological abilities.....	223
4.6.1.4	Predictors of Performance on the ELO-L Test Battery.....	225
4.6.2	Arabic LITMUS-SRT	229
4.6.2.1	Predictors of performance on Arabic LITMUS-SRT.....	230

4.6.2.2	Performance on the individual structures of the Arabic LITMUS-SRT	232
4.7	Discussion	235
4.7.1	Standardized HL test battery for Arabic	236
4.7.2	Arabic LITMUS-SRT	238
4.7.3	HL abilities of bilingual children with SLI.....	240
4.8	Conclusion	241
5.	Study 2: Influence of Language Dominance on LITMUS-Repetition tasks	243
5.1	Introduction.....	244
5.2	The present study	247
5.3	Method	249
5.3.1	Establishing language dominance in bilingual children	249
5.3.2	Materials and assessment tools	250
5.3.3	Participants	250
5.3.4	Data analyses and statistical treatment	253
5.4	Results.....	258
5.4.1	Overall results on the German LITMUS-QU-NWRT and SRT: Language dominance as a factor influencing performance on LITMUS repetition tasks	258
5.4.2	Results of PAM-cluster analyses	261
5.5	Discussion and conclusion	270
6.	Study 3: German LITMUS sentence-repetition task as a screening tool for bilingual children in Germany: The role of syntactic complexity and working memory	277
6.1	Introduction.....	279
6.1.1	The present study	283
6.2	Method	285
6.2.1	Materials	285
6.2.2	Participants	285
6.2.3	Data analyses and statistical treatment	286
6.3	Results.....	288
6.3.1	Overall performance on the German LITMUS SRT and diagnostic accuracy.....	288
6.3.2	Effects of syntactic complexity	291
6.3.3	Factors predicting performance on German LITMUS-SRT	297
6.4	Discussion	300
6.5	Conclusion	304
7.	General discussion	306
7.1	Summary	306
7.2	Study 1: Evaluating the diagnostic value of heritage-language assessment for ruling in or out SLI in heritage bilinguals (Aim 1)	310

7.3	Study 2: Examining the diagnostic robustness of the German LITMUS sentence and nonword repetition tasks against the influence of language dominance (Aim 2)	313
7.4	Study 3: Examining the differential effects of syntactic complexity on SRT performance in monolingual and bilingual children with and without SLI, and their interaction with language and working-memory abilities (Aim 3)	315
7.5	Conclusion and outlook	319
References		325
Appendices		414
I	Items of the German LITMUS-SRT	414
II	Items of the German LITMUS-QU-NWR	417
III	The LITMUS questionnaire for parents of bilingual children (German version)	418
IV	BiLaD testing protocol (first page)	425
Eidesstattliche Erklärung		426

List of Tables

Table 1: Criteria for SLI	78
Table 2: Overview of standardized L1 and L2 tests	158
Table 3: German LITMUS-SRT: an overview of test conditions	170
Table 4: German LITMUS-QU-NWRT: an overview of segments and syllable types	177
Table 5: Participants: an overview of background variables (Mean, SD, and range) ..	207
Table 6: Arabic LITMUS-SRT: an overview of test conditions (adapted from Saad & Henry, 2020)	211
Table 7: Structures included in the Arabic LITMUS-SRT with examples and glosses	213
Table 8: Group means (SD) on ELO-L subtests (raw scores, percentage correct)	218
Table 9: Group means (SD) on ELO-L subtests (standardized z-scores)	219
Table 10: Pairwise Mann-Whitney U comparisons: ELO-L (percentage correct and z-scores), subtests Receptive (LexR) and Expressive vocabulary (LexP)	220
Table 11: Pairwise Mann-Whitney U comparisons: ELO-L (percentage correct and z-scores), subtests Morphosyntax Comprehension (MorphR) and Morphosyntax Production (MorphP)	222
Table 12: BiTD children: Spearman correlations between ELO-L subtests and PaBiQ age and input variables	225
Table 13: Summary of hierarchical regression analysis for age and input variables predicting performance in the Receptive Vocabulary subtest of the ELO-L	226
Table 14: Summary of hierarchical regression analysis for age and input variables predicting performance in the Expressive Vocabulary subtest of the ELO-L	227
Table 15: Summary of hierarchical regression analysis for age and input variables predicting performance in the Morphosyntax Production subtest of the ELO-L	228
Table 16: Summary of hierarchical regression analysis for age and input variables predicting performance in the Phonology subtest of the ELO-L	229
Table 17: BiTD children: Spearman correlations between Arabic LITMUS-SRT and PaBiQ age and input variables, as well as working memory and linguistic measures	231
Table 18: BiTD children: Summary of hierarchical regression analysis for working memory and linguistic variables	232
Table 19: The effect of group membership on percentage identical repetition on each syntactic structure (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests)	233
Table 20: The effect of group membership on percentage correct target structure on each syntactic structure (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests)	234
Table 21: Participants, including monolingual controls, after verification of clinical status (Mean, SD, and range)	252
Table 22: The effect of (clinical) group membership on LITMUS-QU-NWRT and SRT (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests)	260

Table 23: Participants including monolingual control groups after verification of clinical status: mean (SD) and range	286
Table 24: Syntactic complexity levels based on movement, embedding, and intervention	288
Table 25: Diagnostic accuracy of the German LITMUS-SRT in monolingual and bilingual children: identical repetition (SRT_Id) vs. correct target structure (SRT_Tar)	291
Table 26: Mean (%) SRT_Tar and standard deviation by complexity level	292
Table 27: The effect of (clinical) group membership on different complexity levels of the SRT and post hoc pairwise comparisons	293
Table 28: Summary of hierarchical regression analysis for linguistic and WM factors predicting the performance of BiTD children on the SRT.....	298
Table 29: Summary of hierarchical regression analysis for linguistic and WM factors predicting the performance of BiTD children on Level 1 structures of the German LITMUS SRT	299
Table 30: Summary of hierarchical regression analysis for linguistic and WM factors predicting the performance of MoSLI children on German LITMUS SRT	300

List of Figures

Figure 1: WM components (adopted from Baddeley et al., 2020)	70
Figure 2: PowerPoint Slide from the German LITMUS-SRT (first row)	171
Figure 3: Representation of word-level complexity in a trisyllabic nonword (adapted from Grimm & Hübner, forthcoming). Abbreviations: PW: prosodic word, F: foot, σ : syllable	174
Figure 4: PowerPoint Slide from the German LITMUS-QU-NWRT	177
Figure 5: PaBiQ: Calculation of language exposure index	180
Figure 6: ELO-L (raw scores, %correct): subtests Receptive (LexR) and Expressive Vocabulary (LexP)	220
Figure 7: ELO-L (z-scores): subtests Receptive (LexR) and Expressive Vocabulary (LexP)	221
Figure 8: ELO-L (raw scores, %correct): subtests Morphosyntax Comprehension (MorphR) and Production (MorphP)	222
Figure 9: ELO-L (z-scores): subtests morphosyntax comprehension (MorphR) and production (MorphP)	223
Figure 10: ELO-L: subtest Phonology (raw scores, %correct)	224
Figure 11: ELO-L: subtest Phonology (z-scores)	224
Figure 12: Arabic LITMUS-SRT: % correct identical repetition & target structure ...	230
Figure 13: Arabic LITMUS-SRT: mean percentage of identical repetition on each structure subtype	233
Figure 14: Arabic LITMUS-SRT: Mean percentage of correct target structure on each structure subtype	234
Figure 15: LITMUS-QU-NWRT: percentage of identical repetition: NWRT_global (overall score), NWRT_LI (language-independent part), and NWRT_LD (language-dependent part)	259
Figure 16: SRT: percentage of identical repetition (SRT_Id) and correct production of target structure (SRT_Tar).	259
Figure 17: PAM-clustering of MoSLI+MoTD based on performance scores on SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD	263
Figure 18: PAM-clustering of MoSLI+BiTD based on performance scores on SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD	264
Figure 19: Conditional Inference Tree Analysis of the MoSLI + BiTD clustering result using SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD as predictors	266
Figure 20: Conditional Inference Tree Analysis of the BiSLI+BiTD clustering result based on SRT_Id and SRT_Tar: hierarchical variable importance of background variables predicting cluster membership	267
Figure 21: PAM-clustering of BiSLI+BiTD based on performance scores on SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD	268

Figure 22: Conditional Inference Tree Analysis of the BiSLI+BiTD clustering result using SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD as predictors	269
Figure 23: PAM-clustering of BiSLI+BiTD using performance scores on SRT_Tar and NWRT_LD.....	269
Figure 24: German LITMUS-SRT: Percentage identical repetition and correct target structure.....	290
Figure 25: ROC curves for German LITMUS-SRT: Identical repetition and correct target structure for monolingual (left) and bilingual (right) children.....	290
Figure 26: % SRT_Tar for sentences with and without computational complexity factors	292
Figure 27: Complexity Level 1: Percentage of sentences with correct target structure	294
Figure 28: Complexity Levels 2 & 3: Percentage of sentences with correct target structure	295
Figure 29: Complexity Level 1: Percentage of sentences with correct target structure (BiTDs split by age).....	296
Figure 30: Complexity Level 2: Percentage of sentences with correct target structure (BiTDs split by age).....	296

1 General introduction

Linguistic research on the intersection between bilingualism and language impairment has become highly relevant in light of the ongoing global demographic changes and the accompanying cultural and linguistic diversity. Currently, about 21.2¹ million people (ca. 25% of the overall population) residing in Germany have a migratory background (Federal Statistical Office [Destatis], 2019), and the trend is on the rise. The proportion of preschool children who speak languages other than German at home is approximately 40.6%, and in a number of urban areas, e.g., in Hamburg and Berlin, the figure is higher than 50% (Bundeszentrale für politische Bildung, 2021).

Although speaking more than one language neither causes nor puts the child at an increased risk for developing language and communication disorders (Genesee, 2015; Kohnert et al., 2020; Paradis, 2010a, 2010b; Rothweiler, 2013), bilingual children are frequently under- and overrepresented in clinical caseloads (Bedore & Peña, 2008; Crutchley et al., 1997; de Jong et al., 2010; Dockrell et al., 2014; Grimm & Schulz, 2014; Peña et al., 2011; Salameh et al., 2002; Scharff-Rethfeldt, 2013; Smeets et al., 2009; Stow & Dodd, 2005; Williams & McLeod, 2012). Underdiagnosis (also known as “missed identity”) occurs when language delays/deficits are overlooked and misinterpreted as a natural consequence of emerging bilingualism (Bedore & Peña, 2008; Genesee et al., 2004). Conversely, overdiagnosis (also referred to as “mistaken identity”) occurs if interlanguage features are mistakenly identified as a sign of language impairment when insufficient exposure to the language of assessment is responsible for the child’s temporary language deficits. In the case of underdiagnosis, a bilingual child with genuine underlying language impairment is not given timely access to appropriate therapeutic language intervention, which is crucial for ameliorating long-term problems associated with the impairment and ensuring optimal outcomes. In the latter case scenario, a typically developing bilingual child (BiTD), who merely needs language support, is unnecessarily referred to speech and language therapy services. As a result, “limited resources may be misappropriated; parents, family members, and children are needlessly concerned; or attention is refocused from other developmentally appropriate activities” (Kohnert, 2013, p. 146). Not only is this costly for the healthcare system, but it also bears a risk of

¹ This figure includes residents born abroad as well subsequent generations born in Germany.

stigmatization for the child (Neumann et al., 2024; Schulz et al., 2017; Zurer-Pearson, 2010).

Recent research on language development has focused on problems of language assessment and the identification of Specific Language Impairment (SLI), or what is also referred to as Developmental Language Disorder (DLD)² in the current literature (see Bishop et al., 2017, and Rice, 2020a for discussions), in bilingual contexts. A major source of diagnostic difficulties among monolinguals is the substantial heterogeneity of the linguistic profiles of children with SLI, who constitute a group with diverse profiles and deficits varying in severity across language components (Conti-Ramsden et al., 2001a; Crutchley et al., 1997; Friedmann & Novogrodsky, 2011; Leonard, 2014a, among others). Apart from diagnostic challenges caused by the heterogeneity of the disorder itself, identifying language impairment in bilingual children is made far more complex by the great variability in their (typical) language development, which is shaped by a myriad of child internal and external factors (Birdsong, 2018; Grüter & Paradis, 2014; Hamann, 2012; Paradis, 2023). The latter include inter alia, timing of bilingualism, length of exposure to each of the languages, quantity and quality of linguistic input (poor or enriched), typological proximity of the languages in question, the status of the home language (high prestige, minority or heritage language), and the socioeconomic status of the family (see Czinglar et al., 2015 and Hoff, 2006 for overviews). The interplay of these factors makes it notoriously difficult to establish what is typical for bilingual language development (Tuller et al., 2018). Assessment is further complicated by the fact that the developmental trajectory of bilingual child acquisition may show (persistent) delays or temporary overlap with that of monolingual children with SLI (MoSLI), especially in the area of morphosyntax (Chilla, 2008; de Jong et al., 2010; Håkansson & Nettelbladt, 1993; Hamann, 2012; Paradis & Crago, 2004; Paradis et al., 2010; Paradis et al., 2016; Rothweiler, 2006; Rothweiler et al., 2012; Tuller et al., 2015, and many others). In the case of unimpaired bilinguals, the temporary overlap is very likely to gradually resolve with more exposure to the ambient language (Hamann & Belletti, 2006; Paradis, 2010a, 2010b, 2010c). However, studies show that child bilinguals can take more than three years

² The author is aware of the recent preference for using the term “Developmental Language Disorder-DLD” for unexplained language impairment in the absence of primary deficits. However, the label “SLI” is used throughout this thesis for the sake of continuity with much of the existing literature on bilingual SLI and own collaborative research (see also Rice, 2020a for a discussion). Moreover, the sample characteristics in the studies presented in this thesis are more consistent with the diagnostic label “SLI”, since only children with a non-verbal IQ of 85 and above were included in the studies.

of exposure to the second language (L2) to converge with their monolingual peers for most linguistic subdomains and show individual variation depending on their language environment factors (Blom et al., 2012; Bohman et al., 2010; Paradis et al., 2016). Especially in the case of sequential bilinguals, language screening of preschool children takes place during a stage of bilingualism in which their abilities in the societal language are still emerging (Paradis, 2016).

Since genuine language impairment affects any language that is being acquired, a bilingual child should be ideally assessed in both of her languages (ASHA, 2023; IALP, 2011; RCSLT, 2007) or at least in her first language (L1, i.e., home/heritage language) if it is the dominant one (Fredman, 2006). However, L1 assessment is not always feasible due to the lack of L1-speaking speech-language pathologists (SLPs) and adequate norm-referenced tests for bilingual children acquiring their L1 in heritage contexts, i.e., in contexts where the L1 is not the ambient language (Rothman, 2009). This leaves practitioners with no option but to resort to direct assessment measures in the L2. Diagnostic problems mainly occur when bilingual children are solely assessed using monolingual norm-referenced tests in the majority/societal language, which might still be their weaker, i.e., non-dominant language at the time of assessment. In many cases, performance below the monolingual average, especially on standardized measures for vocabulary and morphosyntax, is taken as evidence for language impairment leading to overdiagnosis with SLI (Bedore & Peña, 2008; Grimm & Schulz, 2014). Yet, postponing referral to speech-language therapy until a bilingual child has had sufficient exposure to the L2 would mean that a bilingual child with genuine language impairment is not referred to differential diagnosis until it is too late to optimally profit from speech-language intervention (Wiefferink et al., 2020), which is not that uncommon in the case of sequential bilinguals with limited exposure to the L2 (cf. Nayeb et al., 2015; Salameh et al., 2002; Voet Cornelli, 2020). In light of the aforementioned diagnostic problems, clinical practitioners and educators constantly face the challenge of how to determine whether a bilingual child's poor performance on language measures in the ambient language (second language, L2) is the result of an inborn language impairment or is due to insufficient exposure to the L2, which is typically the language of assessment (cf. Armon-Lotem, 2018; Armon-Lotem et al., 2015a; Hamann, 2012; Marinis et al., 2017; Tuller et al., 2015, 2018).

The paucity of appropriate assessment tools for bilingual children and the arising diagnostic difficulties are highly relevant for the German context since the German school

system provides special support for children with language deficits and hence requires language assessment of all children at risk of language impairment (Deml et al., 2018; see also Hamann et al., 2020 and Hertel et al., 2022). In Germany, all children aged between five and six undergo a statutory screening (school entry examination; *Schuleingangsuntersuchung*) prior to school enrollment. The examination is usually performed by public youth health practitioners or physicians appointed by the regional Department of Public Health and assesses aspects relevant to academic achievement in school. The latter include motor skills, cognitive and language abilities, and socio-emotional competencies. Within this screening, direct language assessment is conducted using standardized or informal tests and via evaluation of children's spontaneous speech samples (Voet Cornelli et al., 2012). Notably, the employed procedures are not uniform and might differ at federal-state and regional levels (Kelle, 2011; Scharff Rethfeldt, 2019). A close inspection of the published annual reports for some of the states with high proportions of children with migration background, such as Berlin (Bettge & Oberwöhrmann, 2018), Hamburg (Heckt & Pohlmann, 2018), Lower Saxony (Bruns-Philipps et al., 2018), and Bremen (Tempel et al., 2018), reveals that bilingual children score disproportionately low on the subtests assessing German language skills. For example, according to Berlin's report of preschool evaluation for the year 2017 (Bettge & Oberwöhrmann, 2018), language deficits were observed in more than 50% of the assessed bilingual children, as opposed to only 8.7% among monolinguals. In particular, preschool children of parents with low educational levels, who are primarily exposed to the minority language in the home environment combined with a late kindergarten start, are being "increasingly diagnosed with linguistic deficits" in the majority language and are much more likely to receive a recommendation for delayed school-entry as well as further differential diagnostics than their monolingual German-speaking peers (Authoring Group Educational Reporting, 2016, p.14; Tuppat et al., 2016; see Nayeb et al., 2015 for similar results in Sweden). Thus, there seems to be a link between migration status and socioeconomic status and the migration-specific disadvantages in acquiring school-related skills, among which are language skills (Becker & Beck, 2011; Klieme et al., 2010; Powell & Wagner, 2014). In the same vein, the reports of one of Germany's most prominent public insurance companies, AOK, state that one in four six-year-old bilingual children is diagnosed with language impairment (Waltersbacher, 2016). These results are alarming from a public health perspective since these figures are substantially higher than the estimated 7.4% language impairment prevalence rate among monolingual

and bilingual preschool children (Tomblin et al., 1997; Norbury et al., 2016) and reflect a growing trend for medicalizing emerging bilingualism.

In order to enhance diagnostic practices, clinical and educational professionals would need to have an in-depth understanding of what constitutes typical language development in bilingual child language acquisition and how such children compare to their language-impaired peers. This also includes knowing the expected trajectories of both simultaneous and sequential bilingual language development as a function of the age of acquisition and amount of exposure (see Haberzettl, 2005, 2014; Kohnert, 2013). Moreover, speech-language pathologists (SLPs) would greatly benefit from having “evidence-based alternatives to monolingual assessment tools and protocols for [bilingual] clients” (Paradis, 2016, p. 171). However, in clinical reality, appropriate standardized tools for the assessment of child bilinguals are scarce, and a considerable number of clinical practitioners lack the necessary theoretical and practical training in working with bilingual children with language impairments (Babatsouli, 2021; Guiberson & Atkins, 2012; Newbury et al., 2020). For example, a cross-sectional survey conducted by Scharff Rethfeldt (2019) in the state of Bremen in Germany revealed that more than 40% of the participating SLPs did not receive additional training on working with multilingual³ children. Hence, given the decisive role of early identification and targeted intervention for ameliorating the long-term outcomes of SLI, there is an urgent need for reliable assessment tools for “ascertaining or excluding [language impairment] that do not involve waiting until sufficient exposure would (presumably) allow for resolution of language difficulties, in the case of” bilingual children with typical language development (Tuller et al., 2015). Ideally, the tools should be fast and easy to administer and score, given the limited time and staff resources typically allocated for language screening.

In a collaborative endeavor to cope with diagnostic challenges in multilingual contexts, the COST Action IS0804 ‘*Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment*’ (<https://www.cost.eu/actions/IS0804/>) was set up by the European Cooperation in Science and Technology (COST: <https://www.cost.eu/>). It brought together researchers and clinical practitioners from over 24 countries with the goal of profiling bilingual specific language impairment (BiSLI, see www.bi-sli.org) and developing a comprehensive set of linguistic and cognitive tools for

³ Note that the terms ‘bilingual’ and ‘multilingual’ are used interchangeably to refer to children acquiring two or more languages.

identifying SLI in bilingual children across many languages. The latter tools are known as the LITMUS tools (**L**anguage **I**mpairment **T**esting in **M**ultilingual **S**ettings) and were constructed according to a set of principles that allow cross-linguistic comparability (see Armon-Lotem et al., 2015a). Two prominent linguistic LITMUS tools of particular relevance for the current thesis are sentence repetition (SRT) and nonword repetition tasks (NWRT), which are known to have excellent diagnostic accuracy in monolinguals (Conti-Ramsden et al., 2001a). SRT taps morpho-syntactic competence (Chiat et al., 2013; Fleckstein et al., 2018; Hamann et al., 2017; Polišenská et al., 2015; Vinther, 2002), while NWRT, when constructed as such, taps phonological abilities (Chiat, 2015; Chiat & Polišenská, 2016; dos Santos & Ferré, 2018; Grimm, 2022). The latter two domains are reported to be particularly affected in individuals with SLI (Leonard, 2014a). Importantly, instead of only focusing on language-specific vulnerable domains that could penalize bilingual children, LITMUS SRTs and NWRTs primarily incorporate (morpho)syntactic and phonological constructions known to be cross-linguistically challenging for language-impaired children but not necessarily for bilingual typically developing children. Examples of such structures are aspects of computational (e.g., derivationally complex constructions containing *wh*-movement, clausal embedding, and intervention, Friedmann et al., 2009; Hamann et al., 2007; Tuller et al., 2012) and phonological complexity (e.g., presence of consonant clusters and consonantal coda, Gallon et al., 2007; Ferré et al., 2015). As such, the tasks should be able to identify language impairment in bilingual children without disadvantaging them (Armon-Lotem et al., 2015a). Results of many recent studies probing LITMUS tools, especially LITMUS SRTs and NWRTs, with bilingual populations are very encouraging. The tasks appear to reliably distinguish between BiTD and BiSLI with different language combinations, especially when used in conjunction (e.g., Armon-Lotem & Meir, 2016; Chilla et al., 2021; de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Komeili & Marshall, 2013; Tuller et al., 2015, 2018; see also Grohmann & Armon-Lotem, 2021 and publications therein). This also applies to the German LITMUS sentence and quasi-universal nonword repetition tasks (e.g., Grimm, 2022; Hamann & Abed Ibrahim, 2017). Several studies further sought to examine the extent to which factors inherent to bilingualism, as well as factors related to the construction of the tasks, e.g., phonological and structural complexity, influence performance on the tasks, especially when administered in the majority language (Armon-Lotem & Meir, 2016; Boerma et al., 2015; de Almeida et al., 2019; Grimm & Schulz, 2021; Hamann et al., 2017; Meir et al., 2016;

Tuller et al., 2018; Zebib et al., 2020b). However, none of the existing studies has examined whether factors related to bilingualism could influence performance on LITMUS repetition tasks to an extent that would compromise their diagnostic accuracy for SLI in bilingual contexts. Moreover, existing studies vary regarding the standard against which diagnostic accuracy was estimated, as there is ‘no agreed-upon standard for detecting language impairment’ for bilingual dual language learners (Tuller et al., 2018, p.3). For instance, while some studies verify clinical status using a strict assessment paradigm in both of the child’s languages alongside parental information (e.g., Armon-Lotem & Meir, 2016; Chilla et al., 2021; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018), other studies base clinical status on referral to SLP services or solely rely on L2 assessment for establishing clinical status (e.g., Boerma et al., 2015; Grimm, 2022; Grimm & Schulz, 2021), sometimes combined with information about the child’s early developmental history and family history of language impairment (e.g., Boerma & Blom, 2017). Moreover, studies on the clinical utility of the LITMUS tools for the assessment of the bilingual child’s L1, which is typically acquired in a minority context, are relatively rare (see Scheidnes, 2020).

1.1 Outline of this dissertation

This dissertation is located at the intersection of bilingualism and SLI and arose within the Franco-German cooperation research project *Bilingual Language Development: Typically Developing Children and Children with Language Impairment* (BiLaD: <https://uol.de/en/cornelia-hamann/research/bilad>). The project's main objective was to define crosslinguistic criteria for identifying SLI in bilingual children regardless of their respective L1-L2 constellations and gain insights into the underlying nature of bilingual SLI. The project further aimed to validate some of the newly developed cross-linguistic LITMUS tools for the German and French contexts that differ in their sociolinguistic settings and provide implications relevant for teacher training and clinical evaluation by speech-language therapists.

The thesis consists of three thematically related empirical studies. Study 1 builds on the recommendation of assessing both languages or at least the dominant language of a bilingual child to avoid cases of misdiagnosis in multilingual contexts, which can play a decisive role, especially in the case of recently arrived bilingual children with limited

exposure to the ambient language (Fredman, 2006). Focusing on Arabic-German bilingual children, the study examines the clinical utility of employing a standardized L1-test battery, namely the ELO-L (Zebib et al., 2017, 2019), for ruling in/out language impairment in bilingual children acquiring Arabic as a heritage language in an L2-German majority context. Given that LITMUS sentence repetition tasks have shown promising results for disentangling impairment from typical language development in bilingual children when administered in the majority language, the study further investigates whether a LITMUS sentence repetition task for Levantine Arabic (Zebib et al., 2020a) can reliably distinguish BiTD from BiSLI in heritage contexts, i.e., in contexts where the L1 is acquired as a minority language. Study 2 focuses on L2 assessment using the newly developed German LITMUS sentence (Hamann et al., 2013) and quasi-universal nonword repetition tasks (Grimm et al., 2014). It specifically addresses the issue of language dominance and examines whether background variables related to the bilingual experience, particularly the degree of language dominance as measured by the relative amount of use and exposure to each of the child's languages, can compromise the diagnostic accuracy of the German LITMUS-SRT and QU-NWRT. Finally, since syntactic complexity (operationalized by the presence of syntactic embedding, verb movement, and intervention) played a central role in the construction of the German LITMUS-SRT, given its vulnerability in SLI and relative robustness in bilingual acquisition, Study 3 investigates how syntactic complexity interacts with repetition accuracy in the different lingual (monolingual vs. bilingual) and clinical groups (TD vs. SLI) and examines whether variance of performance is primarily related to language abilities (Klem et al., 2015) or to working memory capacities as estimated by simple and complex digit recall tasks.

The dissertation is organized as follows: Chapter 2 provides an overview of relevant theoretical background on bilingualism, SLI, and diagnostic challenges in bilingual contexts. Chapter 3 presents the materials and methods used in the empirical studies of this thesis. Chapter 4 is dedicated to Study 1 and includes relevant theoretical background on heritage language development not introduced in Chapter 2. Chapters 5 and 6 present Studies 2 and 3, respectively. Finally, Chapter 7 provides a summary and a discussion of the main findings of the empirical studies with clinical implications and directions for future research.

2 Theoretical background

As an introduction to the studies presented in this thesis, the subsequent sections will provide a general overview of bilingual child language acquisition, specific language impairment, and the intersection of the two. Subsequently, it will address the diagnostic challenges in bilingual contexts and introduce the reader to the LITMUS tools that were recently developed in an attempt to cope with assessment problems in multilingual settings. Note that the theoretical background relevant to heritage language acquisition is presented separately in Chapter 4.

2.1 Bilingual child language acquisition

In the literature on language acquisition, bilingualism has been defined in multiple ways. For this dissertation, I adopt Grosjean's (1992) definition, according to which bilinguals are "those [individuals] who need and use two or more languages in their everyday lives" (p. 51). A key aspect of this definition is the focus on language use regardless of the relative proficiency in each of the used languages making it suitable even for describing bilingual children with (very) limited language abilities, including children with language impairments, who still need to utilize both of their languages to be able to function in their daily lives (Kay-Raining Bird et al., 2016). This thesis focuses on the so-called "child bilinguals", also referred to as "developing bilinguals", i.e., children exposed to two or more languages in childhood from birth up to middle-childhood⁴ (cf. Unsworth et al., 2014a, p.766), i.e., "during the most dynamic period of communication development" (Kohnert, 2013, p. 86; see also Herschensohn, 2007).

Studies on monolingual language acquisition have shown that variations inherent to the language environment of the child, such as input quantity and quality, as well as the complexity and distributional consistency of certain linguistic structures within a given language, impact the rate of language development (e.g., Ambridge et al., 2015; Hurtado et al., 2008; Huttenlocher et al., 2010; Paradis & Grüter, 2014). Moreover, distal experience factors, such as the family's socioeconomic status (SES), have been shown to impact language growth by modulating the nature of proximal factors, such as the

⁴ For a detailed description of different stages in childhood, see Steinberg et al. (2011, p.5).

richness of language input and amount of exposure (Fernald et al., 2013; Hoff, 2006; Huttenlocher et al., 2010; Locke et al., 2002). A crucial difference between monolingual and bilingual acquisition is that children acquiring two languages are subject to more potential sources of variation in their language environment than their monolingual age-peers (Grüter & Paradis, 2014; Paradis & Jia, 2017; see Armon-Lotem & Meir, 2018 and Paradis, 2019, 2023 for overviews). This, in turn, influences various aspects of bilingual language development both within and across the languages of the child (e.g., Hoff, 2018; Paradis, 2010a; Paradis et al., 2014, 2017a, 2017b; Thomas et al., 2014; Thordardottir, 2019; Unsworth et al., 2014b, 2016a, 2016b).

2.1.1 Sources of individual variation in bilingual language development

Children who are raised in a bilingual setting are exposed to at least two languages: their heritage language (HL), also labeled as first language (L1), home language, minority language, or ethnic language (Montrul, 2016a, 2018; Rothman, 2009), and the societal, ambient or majority language, frequently referred to as second first language in case of simultaneous bilingualism (2L1) and second language (L2) in case of sequential bilingualism. A key characteristic of developing bilingualism is the variability in the bilingual language experience, which, alongside factors known to affect monolingual language acquisition, modulates individual outcomes of language acquisition in each of the child's languages (Kohnert, 2013; Paradis, 2023; Unsworth, 2016a, 2016b). Research on bilingual language acquisition has identified multiple mutually non-exclusive language-level and child-level factors contributing to individual differences in child L1 and L2 acquisition (see Armon-Lotem & Meir, 2019; Chondrogianni, 2018; Montrul, 2016b; Paradis, 2019, 2023; Unsworth, 2016a, 2016b for overviews). Language-level factors refer to linguistic and structural properties of the child's two languages, which provide "a resource for each other in development" (Paradis & Jia, 2017, p. 3) and are partially responsible for the "differential acquisition rates for certain linguistic structures" (Paradis, 2019, p.15). These include frequency/regularity, structural complexity, typological proximity, and timing in monolingual acquisition (Grüter & Paradis, 2014; Paradis, 2023). On the other hand, child-level factors include proximal and distal child-external factors related to the child's language environment and child-internal factors, such as chronological age, age of onset of bilingualism, and cognitive capacities of the child, such as working memory capacities and non-verbal analytical reasoning (Paradis

et al., 2011, 2023). More recently, the child's socioemotional well-being has been explored as a further child-internal factor contributing to individual variation in bilingual language acquisition (see Franck & Delage, 2022; Meir & Armon-Lotem, 2017). Child-level properties, among other things, age of onset of bilingualism, cognitive abilities, input quantity, and quality, as well as distal factors, such as the family's socioeconomic status, play a crucial role in modulating "access to –and uptake of –language-level properties, and thus are determinants of acquisition rates and outcomes" (Paradis, 2019, p.15; see also Farnia & Geva, 2011) regardless of which theoretical framework is adopted to investigate child language acquisition (cf. Rankin & Unsworth, 2016; Unsworth 2016b for discussions). In the following sections, I will elaborate on frequently discussed sources of individual variation that are of direct relevance to the current thesis.

2.1.1.1 Child-internal factors

Age of onset of bilingualism

One of the most extensively investigated topics in research on bilingual language acquisition is the effects of the age of onset of exposure to the L2 (AoO, also known as age of acquisition or age of bilingualism) on bilingual language development (Bialystok, 1997; Birdsong, 2018; DeKeyser, 2012; Newport, 1990; Unsworth, 2013c). AoO refers to the age at which systematic and sustained exposure to the L2 begins. It "marks [the] transition from monolingual input to bilingual input" (Altman et al., 2022, p. 1135), which is typically accompanied by a significant change in the quantity and quality of linguistic input in each of the languages. Moreover, AoO is indicative of the level of cognitive maturity at the time bilingualism begins, where a higher age of onset is associated with higher cognitive or (conceptual) maturity (Chondrogianni & Marinis, 2011).

There is a general consensus on the presence of age effects in bilingual language acquisition. When the entire lifespan and ultimate attainment⁵ are considered, an earlier AoO is often viewed as advantageous, at least for some linguistic domains⁶ (see the widely cited study by Johnson & Newport, 1989). For instance, adult L2 learners who

⁵ This thesis does not address (L2) ultimate attainment (final-state grammars) since it is concerned with emergent bilingualism.

⁶ This is not to say that native-like proficiency is not attainable in adult L2 learners. Smithson et al. (2014) as well as White and Genesee (1996) report that monolingual proficiency levels in vocabulary and syntax can be reached in contexts fostering proficiency in the L1 and L2 as in the Canadian sociocultural context for French-English bilingualism (see also Nicoladis, 2018).

start acquiring an L2 post-puberty have often been reported to be “detectably different in speech production from monolingual native speakers and early L2 learners” (Hopp & Schmid, 2013, p. 361; see also Abrahamsson & Hyltenstam, 2009; Flege et al., 2006; Munro & Mann, 2005). More recently, there has been an increased interest in the role AoO plays within childhood bilingualism. A frequently asked question by educators and language professionals is whether similar outcomes are to be expected from bilinguals with varying ages of onset.

Depending on the age of onset of exposure to each of the languages, child bilingualism can be divided into simultaneous bilingualism (also known as bilingual first language acquisition, 2L1; Genesee & Nicoladis, 2007) or sequential, i.e., successive bilingualism (child L2 bilingualism, cL2). In simultaneous bilingual language acquisition, naturalistic acquisition of both languages typically starts during infancy, i.e., mostly at birth or shortly thereafter (cf. De Houwer, 2009, 2021; Meisel, 2004). Despite receiving less input to each of their languages on a day-to-day basis, simultaneous bilingual children have been reported to differentiate⁷ their language systems from the onset of bilingualism and go through the same developmental milestones in each of their languages as their monolingual peers acquiring the respective L1s (Hamann, 2012). Moreover, they have been shown to generally achieve this within the same time frame as their monolingual peers, at least in the area of morphosyntax (see Genesee & Nicoladis, 2007; Hamann, 2012; Meisel, 1994, 2007, 2009 for overviews). Despite qualitative similarities to monolinguals, equal success in both languages is not guaranteed due to the influence of other crucial factors, such as quantity and quality of language input, dominance shift, and the sociolinguistic status of the languages in question. Moreover, quantitative differences in the form of a temporary boost, i.e., acceleration effects (e.g., Kupisch, 2006) or delays in one of the languages (e.g., Eichler et al., 2013; Gathercole & Thomas, 2009) can occur due to crosslinguistic influence, where delays are more pronounced in the lexical domain (Bialystok, 2009; see Hager & Müller, 2015 and Müller, 2017 for discussions).

In successive childhood bilingualism, acquisition of the L2 also starts in childhood, but after considerable progress towards the acquisition of the first language has been made and certain L1 properties have been acquired (Hamann, 2012; Paradis et al., 2021b). Successive bilingualism can be further distinguished, based on how long

⁷ Note that “differentiation does not mean that [2L1 bilinguals] always use both of their languages exactly like monolinguals of each language would” (Nicoladis, 2018, p. 88).

exposure to one language during childhood predates the other, into early (ecL2) and late child L2 acquisition (lcL2). Such distinctions are motivated by age-related alternations in the neuroplasticity of the brain (see Liu et al., 2021 and references therein) leading to the existence of multiple critical or sensitive periods⁸ at which crucial changes in the language-making capacity (LMC⁹), i.e., in the acquisition mechanisms take place (Bhatia & Ritchie, 2013; Birdsong, 2006; Locke, 1997; Meisel, 2004, 2008, 2011, 2013; Slobin, 1985; see Meisel, 2018 for recent discussions). Meisel (2018) defines sensitive periods as “clusters of sensitive phases during which the LMC is optimally prepared to integrate new information into developing grammars” (p.657). At the same time, Meisel stresses that “it is not ‘language’ which is affected by maturational changes but *certain domains of grammar* [Meisel’s emphasis]. The acquisition of lexical knowledge, for example, is not concerned at all” (Meisel, 2013, p.76). Note that the existence of sensitive periods does not entail that learning a second language becomes impossible after a particular age, but rather assumes that learning certain aspects of grammar¹⁰ requires different learning mechanisms.

⁸ This was originally postulated as the Critical Period Hypothesis (CPH) in the seminal work of Lenneberg (1967) according to which maturational effects hindering naturalistic language acquisition emerge around puberty (around age 11). In light of more recent psycholinguistic and neurolinguistic research, CPH was later modified to accommodate the existence of more than one critical period, i.e., sensitive periods (see Locke, 1997; Long, 1990; Meisel, 2013).

⁹ Note that the LMC comprises both domain-specific and domain-general mechanisms (Meisel, 2018, p. 662).

¹⁰ Different proposals have been put forward to characterize the nature of such grammatical knowledge, e.g., declarative or procedural representations that are the result of explicit versus implicit learning (e.g., M. Paradis, 2009). Some researchers argue that (adult) L2 learners employ alternative strategies drawing on their domain-general problem-solving skills as coping mechanisms to compensate for underlying representational deficits (see DeKeyser, 2000; Hawkins, 2001; Hawkins & Hattori, 2006), which arise “whenever the L1 and L2 differ in terms of the features that are acquired” (White, 2018, p.65), especially in case of uninterpretable features or parameterized principles of Universal Grammar (Hamann, 2000; Smith & Tsimpli, 1995; White, 1991). This implies that even in cases where adult L2 learners demonstrate L2 native-like levels, especially in “offline tasks that allow the application of explicit knowledge”, there is a fundamental “difference in the underlying nature of [their] linguistic knowledge” (Schmid et al., 2014, p. 130, see also Abrahamsson & Hyltenstam, 2009; Tsimpli & Dimitrakopoulou, 2007). Against the view that representations of certain features in the interlanguage of L2 learners are defectively or only partially represented, the Missing Surface Inflection Hypothesis (Lardiere, 1998, 2000; Prévost & White, 2000) assumes that the representations of certain features (representations of functional categories) can be available in the learner’s mental grammar even when L2 learners lack them in production. Under this view deficits in performance reflect difficulties mapping the underlying structure onto the language-specific surface form due to “procedural constraints or competition from the L1” (Schmid et al., 2014, p. 134; see also White, 2018).

There is no consensus concerning the exact age of onset of dual exposure, which demarcates simultaneous from successive bilingual language acquisition. This is partly because varying acquisition ages may differentially influence different aspects of language, e.g., early vs. late acquired phenomena (Nicoladis, 2018, p.82; see also Long, 1990; Singleton, 2001 for discussions). Different cutoffs have been proposed based on studies examining whether “differences in linguistic competence in adulthood can be traced to differences in age of acquisition within preschool years” (Paradis et al., 2021a, p.8) and on studies contrasting the performance of child L2 bilinguals with varying AoO to that of their 2L1 peers and adult L2 (aL2) learners. (e.g., Hulk & Cornips, 2006; Kroffke et al., 2007; Rothweiler, 2006; Schulz & Grimm, 2019; Schwartz, 2004; Unsworth et al., 2014a). According to Long (1990), crucial changes occur around age 6-7. In her review article, Paradis (2007a) reported that an onset of bilingualism after age 7 resulted in reduced performance on grammaticality judgment tasks (e.g., Weber-Fox & Neville, 1999, 2001). However, subsequent neurolinguistic and behavioral studies showed that changes influencing the acquisition of morphosyntax already happen before age 4 (see Meisel, 2004, 2009). In recent literature, child bilinguals have frequently been classified into early successive child bilinguals¹¹ (ecL2) when AoO lies between 3 and 4 years of age, and late successive child bilinguals, whose first exposure to the L2 takes place between ages 5 and 8 (Hamann, 2012).

Based on longitudinal data on the acquisition of grammatical gender in French¹² by bilingual German-French children (AoO 2;8-4;0) acquiring French in an immersion setting in Germany, Meisel (2008, 2009, 2018) identified the age of 3;6 as a cutoff separating simultaneous from successive child bilingualism. His results showed that children exposed to French before the age of 3;6 behaved rather like monolingual French-speaking children, whereas children with an AoO > 3;6 resembled adult L2 learners in that they mainly relied on lexical learning instead of relying on formal cues, such as the distributional and phonological properties of French nouns. Not only did the children with an AoO > 3;6 demonstrate higher rates of errors, but they also displayed unique error patterns not attested in monolingual acquisition. For instance, they produced gender errors that go against the phonological properties of French nouns, e.g., **la éléphant* ‘the-

¹¹ Note that researchers might use a different cutoff to mark early successive child bilingualism. For example, for Schulz & Tracy (2011) children with AoO $\geq 2;0$ constitute ecL2 learners, while Unsworth et al. (2014a) consider children between ages 1 and 4 to be early successive child bilinguals.

¹² This represents an interface phenomenon between morphology and syntax (Meisel, 2018).

fem. elephant’ or fluctuated between masculine and feminine gender, e.g., *un/la maison* ‘a masc./the-fem. house’, and used incorrect gender markings on resumptive pronouns, e.g., *la fenêtre...il est cassé* ‘the-fem window, he is broken’. Meisel argues that such error patterns “do not reflect knowledge of the target grammatical system” and are most probably the result of lexical item-by-item learning (Meisel, 2018, p. 666f). This suggests that at least for some morphosyntactic phenomena, children with AoO > 3;6 cannot make use of the discovery principles on which monolingual and 2L1 children rely and have to resort to general learning compensatory mechanisms, such as item-by-item learning; see Meisel (2008) for similar results on subject-verb-agreement in French; for German see Chilla (2008), Haberzettl (2005) and Rothweiler (2006), but see Unsworth et al. (2014a) for contrasting results on the acquisition of Dutch neuter gender. Concerning syntax, similar findings were reported for object clitic placement in L2 French (Granfeldt et al., 2007; Meisel, 2008; Tuller et al., 2011) and V2 placement in German (Kroffke et al., 2007; Rothweiler, 2006; Sopata, 2011). At the same time, it has been shown that child L2 learners with an AoO between 3;0 and 4;0 resemble monolingual and 2L1 learners in other aspects. For instance, in French L2, subject clitics are acquired before object clitics (e.g., Hamann & Belletti, 2006; Prévost & White, 2000; Tuller et al., 2011). Similarly, despite delayed mastery, early successive child L2 learners have been reported to simultaneously acquire finiteness (subject-verb agreement) and V2 placement in German, i.e., demonstrate the association between syntax and morphology characteristic for (typical) monolingual development (Chilla, 2008; Kroffke & Rothweiler, 2006; Prévost, 2003; Rothweiler et al., 2017; Tracy & Thoma, 2009; Schulz & Schwarze, 2017). For example, Chilla (2008) observed that early successive German-Turkish bilinguals with an AoO of 3 behaved similarly to monolingual and 2L1 children demonstrating an apparent contingency between finiteness and verb placement in German (cf. Poeppel & Wexler, 1993). In contrast, late successive child bilinguals with an AoO of 6 years and above exhibited aL2 error patterns in that they produced nonfinite verb forms in V2 positions. In the same vein, Roesch and Chondrogianni (2016) found that ecL2 children with AoO around 3;0 (AoO range: 2;9-3;4 yrs.) patterned similarly with their simultaneous French-German bilingual peers on comprehension of *wh*-questions in German in terms of error patterns. Taken together, children with AoO between 3;0 and 4;0 appear to resemble 2L1 learners in some aspects of language acquisition and adult L2 learners in others, making them a distinct acquisitional type within childhood bilingualism.

It is important to note that the influence of AoO on the acquisition process is not across the board, as it interacts with the timing of acquisition of the respective linguistic structures in monolingual acquisition ‘timing in L1 acquisition’, i.e., the age at which a particular structure is acquired by monolingual children (Schulz & Grimm, 2019; Tsimpli, 2014). According to the timing hypothesis, different linguistic structures have different maturational schedules. Some grammatical properties, especially those involving macro-parameters (e.g., word order or *pro*-drop), are argued to require less input to stabilize and are acquired earlier than others in monolingual acquisition, i.e., before the age of four. In contrast, aspects of complex syntax, semantics, pragmatics, and interface phenomena that reside outside “narrow syntax” (e.g., binding properties, subjunctives, structures involving long-distance dependencies, and complex nominals) are usually acquired later in monolingual acquisition since they require more input to be fixed. Late phenomena are acquired around age 5, while very late phenomena are acquired at age six or even later (Schulz & Grimm, 2019). Moreover, late phenomena are assumed to require more cognitive resources due to their relatively higher processing costs. Tsimpli (2014) argues that early acquired phenomena, such as grammatical gender in Greek, “can differentiate between simultaneous and (early) successive bilingualism with an advantage for the former group”. Conversely, late (e.g., sentential negation in German) or very late acquired structures, such as case in German and grammatical gender in Dutch, are expected to “reveal similarly (high or low) performance across bilingual groups, differentiating them from monolinguals” (p. 238). Hence, performance on later acquired linguistic structures may be influenced by factors other than AoO, such as length of exposure or cognitive resources, which often correlate with chronological age (Chondrogianni, 2018; Schulz & Grimm, 2019). Schulz and Grimm (2019) point out that the 2L1-ecL2 differences observed in Meisel’s (2018) study can be explained by the timing hypothesis, as gender represents an early-acquired property in French. This account would also explain the AoO effects found for finiteness and V2 placement in Chilla (2008), which represent early acquired phenomena in German (Clahsen, 1986a; Tracy, 1991, see Schulz, 2007, and Schulz & Grimm, 2012 for overviews on timing in monolingual German acquisition). Timing in acquisition can also account for the similar performance of 2L1 and ecL2 bilinguals on neuter gender marking in Unsworth et al.’s (2014a) study, as it belongs to the very late-acquired linguistic structures in Dutch (Blom et al., 2008). In a more recent study, Schulz and Grimm (2019) investigated timing effects on the acquisition of early (SVA and verb semantics/telicity), late (comprehension of

certain types of *wh*-questions, e.g., adjunct *wh*-questions¹³ and complex sentences), and very late phenomena (negation and case marking) in German by bilingual 2L1 and ecL2 children relative to monolingual German-speaking children. Their results showed that 2L1 learners patterned with German monolinguals outperforming ecL2 children, whereas their performance resembled that of ecL2 children on late and very late acquired phenomena, but see Roesch and Chondrogianni (2016) who investigated the effect of overt case marking as a disambiguating cue on the comprehension of German *wh*-questions in French-German bilinguals. As a consequence, when assessing bilingual children, the ‘timing in L1 acquisition’ of the structures under investigation should be considered in relation to the child’s chronological age because if younger bilingual children (especially when young age is paired with a low LoE) are evaluated on structures that are acquired at a later age in monolingual acquisition, such as case marking (see Eisenbeiss et al., 2006; Lemmer, 2018; Ruigendijk, 2015; Scherger, 2018, 2019a; Schönenberger et al., 2012), “they will inevitably show floor effects” (Grimm & Schulz, 2021, p. 267).

Although an earlier AoO is often associated with superior performance in the L2, especially when ultimate attainment in morphosyntax is considered (Abrahamsson & Hyltenstam, 2009; DeKeyser, 2012; Paradis et al., 2016), an older age of onset of exposure to the L2 within early childhood bears developmental advantages that may have a facilitating effect on the acquisition of certain linguistic domains, such as L2 vocabulary and complex constructions that are associated with high processing demands (Blom & Bosma, 2016; Goldberg et al., 2008; Paradis et al., 2017b; Rothman et al., 2016). This is

¹³ Although production of *wh*-questions starts relatively early in German L1 acquisition, especially that of argument questions (Penner, 1994; Tracy, 1994; see Stromswold, 1995 for English), certain types of *wh*-questions, in particular adjunct questions and questions with overt accusative and dative marking on the *wh*-operator are argued to be mastered relatively late (Schulz, 2013b). Concerning argument questions, object relatives tend to be acquired (in particular comprehension thereof) later than subject relatives (e.g., German: Biran & Ruigendijk, 2015; Italian: De Vincenzi et al., 1999; Hebrew: Friedmann & Szterman, 2011; Palestinian Arabic: Friedmann & Haddad-Hanna, 2014). Moreover, a particular difficulty is reported for referential object *wh*-NP questions (Friedmann et al., 2009, but see Biran & Ruigendijk for contrasting results for German). In their article, Schulz and Grimm (2019) classify *wh*-questions as late acquired phenomena based on Schulz’s (2013b) longitudinal study within the MILA project, in which she examined the comprehension of subject, object and adjunct questions in monolingual and early successive bilingual children acquiring German as L2. She found that comprehension of *wh*-questions gradually improved between ages 3;8 and 5;8. As to the age of acquisition, the results showed that L1 German children mastered subject *wh*-questions at age 3;8 and object *wh*-questions at age 4;8, whereas adjunct *wh*-questions remained problematic at age 5;8. Interestingly, the same pattern was observed for the bilingual group with a lag of 12 months.

ascribed to the fact that children with a later age of L2 onset approach L2 acquisition with greater cognitive (e.g., working memory) and social resources and are linguistically more mature than their 2L1 and ecL2 peers with an earlier AoO (Peets & Bialystok, 2010). For example, bilingual children with an older AoO have been reported to proceed faster than their 2L1 in the acquisition of L2 lexicon (e.g., Blom & Bosma, 2016; Goldberg et al., 2008). In a longitudinal study, Goldberg et al. (2008) followed the lexical development of 19 bilingual children acquiring English as an L2 for three years (mean age at first measuring point = 5;4 [years; months]). They found that children with an AoO > 5;0 had larger vocabularies than children with an AoO < 5;0 at each testing round despite having the same length of exposure to English. In another study, Paradis (2011) reported superior performance of bilingual children (mean age = 5;0) with an older AoO over children with an earlier AoO on measures of L2 lexicon, although both groups were exposed to L2 English for 20 months at the time of assessment. According to Chondrogianni (2018), bilingual children with a later AoO to the L2 have an advantage in L2 lexical acquisition as they can rely on conceptual transfer from L1 to L2. Thus, older L2 children mainly need “to learn new labels for existing concepts already established in their L1” (p. 107). Likewise, an advantage has been observed for the acquisition of morphosyntax, such as verb inflection in English (Paradis et al., 2011) and Dutch (Blom & Baayen, 2013). As to the domain of syntax, Rothman et al. (2016) compared the performance of Spanish-English bilingual children with AoO of 4 to 5 years to child bilinguals first exposed to English between ages 6 and 7 on the comprehension of (long) passive constructions in English. Passives represent a late-acquired structure in English and are argued to be relatively taxing for cognitive processing resources given their computational complexity (Armon-Lotem et al., 2015b). Their results showed that the group with an older AoO reached ceiling performance earlier than the group with an earlier AoO despite being matched on the length of exposure to the L2. A further study by Paradis et al. (2017b) with a sample of 187 English L2 children (mean age = 5;10; [years; months]; mean LoE = 17 months) revealed that even older child L2 learners who had a few months of exposure to English produced a higher proportion of complex constructions than their younger peers. Interestingly, older child L2 learners diverged from acquisitional sequences documented in monolingual acquisition in that they did not show “any order of emergence” for complex sentence types (Paradis et al., 2017b, p. 161). Instead, they simultaneously produced all types of (complex) structures, such as complement clauses, embedded clauses, relative clauses, and coordinated constructions. Paradis (2019) and

Rothman et al. (2016) attribute this discrepancy to the advanced cognitive mechanisms of the children with an older AoO in addition to linguistic experience from the L1, allowing them to acquire such computationally complex constructions at a faster pace.

Cognitive abilities

Cognitive capacities constitute a further factor that is involved in language learning. The latter include inter alia, analytic reasoning skills, working memory capacities (verbal short-term memory and verbal working memory), and executive functions (EFs), such as updating, inhibition, attention allocation, and cognitive flexibility or shifting, i.e., the ability to shift perspectives and switch between sets of rules, tasks and mental states (e.g., Blom, 2019; Gathercole et al., 1992; Verhagen & Leseman, 2016; see 2.1.3 for details). Cognitive capacities are argued to be particularly implicated in the acquisition of complex syntax in monolinguals (e.g., Delage & Frauenfelder, 2019; Engel de Abreu & Gathercole, 2012; Kidd, 2013) and constitute a component part of language aptitude, a predictor of language learning success, especially among adult L2 learners and older children (Harley & Hart, 1997; Miyake & Friedman, 1998; see Doughty, 2019 and Skehan, 2012 for extensive descriptions). Hence, they are also expected to be implicated in bilingual child language acquisition (Paradis, 2023).

Many studies found cognitive abilities, in particular, working memory abilities and nonverbal IQ, to predict individual variation in school-age bilingual children's L2 vocabulary (e.g., Blom, 2019; Lauro et al., 2020), inflectional morphology (Blom et al., 2021b), (complex) syntax and narrative production abilities (e.g., Hamann et al., 2020; Paradis et al., 2011; Paradis et al., 2017b; Sorenson Duncan & Paradis, 2020a; Unsworth et al., 2019). Unsworth et al. (2019) argue that reliance on “domain-general working memory skills” is likely to increase in cases of limited L2 proficiency (p. 1212). This appears to be particularly the case in recently arrived Syrian school-age refugee children investigated by Paradis and her team (Canada), where verbal working memory and analytic reasoning significantly contributed to predicting variance in performance on measures of L2 (English) lexicon, morphosyntax, and inflectional morphology (Paradis et al., 2020, 2021b; Soto-Corominas et al., 2022). Interestingly, some studies showed that cognitive abilities continue to play a role in the development of L2 vocabulary and morphological abilities even after an extended period of exposure to the L2 in majority contexts (e.g., Farnia & Geva, 2011; Paradis et al., 2016; see Paradis, 2023, for an overview).

2.1.1.2 Child-external factors

Language environment factors in bilingual child language acquisition: quantity and quality of L2 input

It is indisputable that bilinguals “have their input space divided” between two languages (Paradis & Genesee, 1996, p. 9) and receive less exposure to each language, on average (both cumulatively and at any given point of time), than monolinguals acquiring the respective languages. In most cases, the input is unevenly distributed between the two languages, both in terms of the amount of time and the domains and social functions in which each of the languages is used. This “distributed characteristic” of bilingual exposure is known as the Complementarity Principle (Grosjean, 1985, 1997; see also Bialystok et al., 2010b; Oller et al., 2007; Patterson & Pearson, 2004). Moreover, the relative amount of input may change over time depending on social and educational settings, i.e., due to changes in patterns of use and exposure resulting from “changes in family structure, child-care arrangements, schooling, or place of residence” (Paradis, 2010c, p. 652). This applies to children who are exposed to both of their languages from birth or shortly after, as well as to those exposed to the second language at a later point in childhood (Paradis & Grüter, 2014). Moreover, the quality of language input bilingual children receive may differ from that of their monolingual peers, not only in the L2 but also in their L1 (Unsworth, 2013a, 2013b, 2016b). Importantly, both the amount and quality of linguistic input bilingual children receive (proximal input factors) are modulated by distal input factors, such as the family’s socioeconomic status, which is frequently operationalized by years of maternal education (Grüter & Paradis, 2014).

Quality of L2 input

“All input is not equal” (Unsworth, 2016b, p.116). In addition to input quantity, numerous studies have demonstrated that qualitative aspects of language input, i.e., the richness of the linguistic environment paired with the family’s socioeconomic status as a distal mediating factor, play an influential role in the development of language abilities of bilingual children, especially at young ages. Linguistic richness refers to the specific characteristics of input in terms of complexity (e.g., frequency of certain linguistic structures), the diversity of input sources, and the density of language-rich activities. Examples of this are TV, literacy-related activities (Duursma et al., 2008) like reading books and storytelling (Quiroz et al., 2010; for results on monolingual children, see

Breitfeld et al., 2021; Demir-Lira et al., 2018; Frijters et al., 2000; Rowe, 2012), interactions with (older) siblings and peers (Sorenson Duncan & Paradis, 2020b), as well as the availability of high-quality input from proficient speakers of the respective language (Unsworth et al., 2019).

A frequent observation is that bilingual children with richer L2 environments demonstrate better L2 skills, e.g., they have a larger vocabulary size, are more accurate on (verbal) morphology, and exhibit an increased use of complex syntactic constructions and more advanced narrative abilities (Govindarajan & Paradis, 2019; Hoff et al., 2014; Jia & Aaronson, 2003; Jia & Fuse, 2007; Paradis et al., 2011; 2017b; Paradis & Kirova, 2014; Prevoo et al., 2014; Scheele et al., 2010). For example, Balilah (2017) examined L1/L2 lexical and overall oral language abilities of 59 Arabic-English bilingual children (age range: 6;0-9;0; mean = 7;11 [years; months]) residing in Ontario (Canada). PPVT-4 (Dunn & Dunn, 2007) and the Expressive Vocabulary Test (EVT; Williams, 2007) were used to assess L2 (English) receptive and expressive vocabulary skills, whereas the CELF-4 test (Semel et al., 2003) was employed for evaluating global performance in L2 English. Results showed that in addition to chronological age and AoO, richness of the English environment outside school significantly contributed towards predicting variance in performance on the CELF-4's composite language score as well as performance on expressive and receptive vocabulary measures, with the explanatory power of L2 richness being more pronounced for the receptive vocabulary measure. In the same vein, Chondrogianni and Marinis (2011) found that children from households with limited literacy activities coupled with mothers' low self-rated L2 proficiency were unable to reach age-appropriate norms on vocabulary and (complex) syntax (e.g., *wh*-questions from the DELV test), see Paradis (2011) for similar results.

Interestingly, Chondrogianni and Marinis (2011) found that the amount of parental L2 input at home did not predict L2 performance, whereas exposure to L2 at school reliably predicted L2 performance (see also Goldberg et al., 2008; Hammer et al., 2014; Paradis et al., 2017b; Paradis & Kirova, 2014; Sorenson Duncan & Paradis, 2019, 2020a, 2020b). Moreover, Paradis' (2011) study revealed that while the child's L2 use at home predicted L2 lexical abilities, parental L2 use at home did not. On the other hand, L2 input from older siblings has been reported to predict L2 performance of younger siblings. Sorenson Duncan and Paradis (2020b) examined the influence of different L2 sources at home, i.e., input from different interlocutors, on L2-English performance of 113 L2 children (age range: 4;10-7;2; mean: 5;10 [years; months]) from different L1

backgrounds. Hierarchical regression modeling, controlling for age, non-verbal IQ, and phonological short-term memory, revealed that better performance in L2 narrative macrostructure, lexicon, and inflectional morphology was associated with L2 input from older siblings but not from mothers. Similarly, Bridges and Hoff (2014) reported superior performance of 16-30-month-old Spanish-English bilingual toddlers with school-aged older siblings on L2-English lexical and grammatical measures. As an explanation for this discrepancy, it has been argued that a significant proportion of the L2 input children receive at home comes from non-native speaker¹⁴ parents, who are often not proficient in the L2 and are thus incapable of providing high-quality L2 input that would promote L2 acquisition (Place & Hoff, 2016; Unsworth et al., 2019; but see Dulay et al., 2017; Nakamura, 2015, and Paradis & Navarro, 2003). Unlike their mothers, older siblings tend to have greater L2 proficiency due to schooling and are thus likely to provide better quality L2 input to their younger siblings. It should be noted that most of the studies above found a strong link between maternal L2 proficiency and educational levels, i.e., the family's socioeconomic status.

Socioeconomic status (SES)

Socioeconomic status SES (as operationalized by levels of education and occupation) represents a broad distal language environment factor that modulates quantitative and qualitative aspects of linguistic input that children receive and, in a sense, serves as an index of “overall family environment and cultural capital” (Paradis et al., 2021b, p. 6; see also Prevoo et al., 2014). In monolingual acquisition, children from low SES backgrounds have been frequently reported to show poorer language outcomes than their age-matched peers from higher SES backgrounds (Rowe, 2018; Rowe & Snow, 2020). In particular, higher levels of maternal education are shown to be associated with more advanced lexical abilities and more complex syntax (e.g., Hoff & Tian, 2005; Hoff et al., 2018a; Huttenlocher et al., 2010; Vernon-Feagans et al., 2019, 2020), presumably because parents with high SES tend to use more varied vocabulary and complex linguistic structures in their interactions with their children (Hart & Risley, 1995; Hoff, 2003a, 2003b), and are thus more likely to provide better-quality enriched language input and to engage in parental activities that foster socioemotional development than caregivers from

¹⁴ In this regard, Armon-Lotem (2018) warns against using dichotomous terms as “native” versus “nonnative” for characterizing input quality and stresses the importance of examining the “actual input directed at children, regardless of preconceived notions of ‘nativeness’” (p. 207).

low SES-backgrounds (Bradley & Corwyn, 2002, McLoyd, 1998; see Hoff, 2006, for an overview). Moreover, several neurocognitive studies found associations between SES and language development in monolingual toddlers (e.g., Benasich et al., 2008, 2014) as well as preschool (e.g., Gou et al., 2011) and school children (e.g., Maguire & Schneider, 2019) as a function of SES-induced changes in brain structure, particularly in areas linked to “memory, executive control, and emotion” (Brito & Noble, 2014, p. 1; see Brito et al., 2016 for an overview). As such, the impact of SES goes beyond merely influencing the quality of parental language input.

As to bilingual children, numerous studies found a connection between the family’s socioeconomic status¹⁵ and performance of bilingual children in the societal language, especially on vocabulary measures (e.g., Armon-Lotem et al., 2011; Calvo & Bialystok, 2014; Cobo-Lewis et al., 2002; Czinglar et al., 2015; Hoff, 2003a; Hoff et al., 2002; Meir & Armon-Lotem, 2017; Oller & Eilers, 2002; Paradis & Jia, 2017). For example, Prevoo et al. (2014) investigated lexical abilities in a sample of 111 bilingual Turkish-Dutch six-year-old children in relation to SES (measured by the family’s income and level of parental education), maternal language use, and reading input at home. They found that the influence of SES on vocabulary in the L2 was mediated by the effect of SES on maternal language use, which in turn modulated home reading input in the L2, i.e., mothers from families with higher SES tended to predominantly use the L2 (Dutch) with their children and to engage more frequently in reading activities in the L2 leading to larger L2 vocabularies. To account for their findings, Prevoo et al. (2014) argued that parents with higher SES are likely to show higher proficiency levels in the L2 and, therefore, can provide better quality input in the L2. A similar observation was made by Sorenson Duncan and Paradis (2020a, 2020b), who reported an indirect link between levels of maternal education and the children’s syntactic abilities in the L2. In a further study, Calvo and Bialystok (2014) compared the performance of 175 6-year-old monolingual and bilingual children with different SES backgrounds (working class vs. middle class) on receptive vocabulary and cognitive measures. As to language measures, their results showed that bilingual children from working-class families had lower receptive lexical skills than working-class monolingual children; at the same time, they were outperformed by monolingual and bilingual children from middle-class families.

¹⁵ In studies on bilingual language acquisition, SES is frequently operationalized as years of maternal education (e.g., Armon-Lotem et al., 2011; Goldberg et al., 2008; Meir & Armon-Lotem, 2017; Paradis & Jia, 2017; Sorenson Duncan & Paradis 2020a, 2020b).

These findings suggest that bilingualism, i.e., split input and low SES, might show cumulative effects, at least for the lexical domain.

In sum, not only proximal input factors but also distal ones like the family's SES should be taken into consideration when assessing bilingual children. This is of great importance given that low SES tends to prevail more among certain groups of immigrants, and might adversely affect overall language development, especially when low-income is paired with low levels of parental education (Andersson et al., 2019; Blom et al., 2019; Fassbender & Leyendecker, 2018; Feyter & Winsler, 2009; Glick et al., 2013), where women are generally affected more than men due to men's higher chances of education (Hartmann, 2016). Additionally, it is important to bear in mind that quantitative and qualitative aspects of the child's language environment are dynamic and may change over time in response to various aspects related to the cognitive and linguistic maturity of the child (De Houwer, 2017). For instance, it has been observed that mothers tend to adjust their interactive behavior according to the child's overall level of language development. For instance, mothers tend to talk more frequently to older children than younger ones (De Houwer, 2014) and are more likely to use more complex language in their interactions with their older children (Poulain & Brauer, 2018). Moreover, a mother might refrain from using a particular language with the child in reaction to the child's weaknesses in that language or the child's preference to only interact in her stronger language (Bonnesen, 2009; De Houwer & Bornstein, 2016b). Such changes in exposure patterns might influence "the overall amount [and quality] of input" a child receives and thus affect linguistic performance in the respective language (De Houwer, 2017, p.19).

2.1.1.3 Cross-linguistic influence (CLI)

A further factor frequently discussed in the literature is cross-linguistic influence (CLI)¹⁶, also known as inter-language influence or transfer. CLI effects within the domains of morphology, core syntax, and its interfaces with other domains have been reported for

¹⁶ Note that CLI here refers to the transfer of linguistic properties from one language into another, and not to the "process by which both languages have access to the same store of knowledge" (Blom et al., 2021c, p. 1162), as originally posited in Cummins (1979) *Interdependence Hypothesis*, which maintains that "on a cognitive level, the languages [of a bilingual child] are not separate but connect to each other by means of common underlying proficiency" (Cummins, 2017, p. 106). Examples of transferable proficiencies include high level abilities that facilitate L2 learning, such as verbal reasoning, metalinguistic awareness (e.g., phonological and morphological awareness) and conceptual knowledge (Bialystok et al., 2014; Chen & Schwartz, 2018; Kroll et al., 2010; Rauch et al., 2012).

adult L2 (see Foley & Flynn, 2013 for an overview) and, more importantly, for simultaneous (Döpke, 1998; Müller & Hulk, 2001; Yip & Mathews, 2002, 2007) and successive child bilingual language acquisition (e.g., Blom & Baayen, 2013; see Chondrogianni & Vasić, 2016 and the publications within their special issue), especially in initial stages of L2 acquisition (e.g., Chondrogianni, 2008; Haznedar, 1997a, 1997b, 2001; Unsworth 2005; Whong-Barr & Schwartz, 2002; Zdorenko & Paradis, 2008). Effects of CLI could manifest quantitatively in the form of acceleration (e.g., Kupisch, 2005, 2007; Tamburelli et al., 2015) or delayed acquisition of certain linguistic properties relative to monolingual acquisition (e.g., Döpke, 1998; Müller & Hulk, 2001; Scherger, 2016). Though less frequent, qualitative manifestations of CLI are attested as well. Unlike quantitative CLI, which only affects the pace of acquisition, qualitative CLI occurs when bilingual children use structures or properties that are unattested in monolingual acquisition under the influence of the other language (see Hulk, 2000, for object clitic placement in French and Schlyter, 1993, for V3 constructions in Swedish).

Various studies have revealed that “CLI does not occur randomly but under well-defined predictable conditions” (Kupisch et al., 2017). Factors proposed to trigger the occurrence of cross-linguistic influence include (partial) surface overlap between the bilingual child’s languages, especially at interfaces (Döpke, 1998; Hulk & Müller, 2000; Müller & Hulk, 2001; Schmitz et al., 2012), language domain, language dominance, as well as markedness and computational complexity (see Kehoe, 2015; Serratrice 2013, for overviews and van Dijk et al., 2022 for a meta-analysis). It is important to note that the latter factors do not operate separately. For example, Müller and Hulk (2001) claimed that surface overlap results in CLI only if it occurs in the C-domain, i.e., “when morphosyntax interacts with discourse pragmatics” as in the case of subject pronouns in null subject languages (e.g., Argyri & Sorace, 2007; Sorace et al., 2009). However, contrary to Hulk, van Dijk et al.’s (2022) meta-analysis revealed that CLI could occur regardless of language domain. Specifically, computationally complex morphosyntactic properties appear susceptible to CLI (Gavarró, 2003; Hopp, 2009; Prévost et al., 2014; Sorace, 2011; Strik & Pérez-Leroux, 2011). Concerning language dominance, it has been argued to influence the directionality of CLI as postulated in, e.g., the *Ivy Hypothesis* (Bernardini & Schlyter, 2004). The *Ivy Hypothesis* maintains that children would transfer functional elements from their dominant language into the non-dominant one because higher functional projections are available to a greater extent in the child’s dominant language. In the same vein, Paradis and Genesee (1996) argue that the developmental

asynchrony¹⁷ that arises due to variations in input leading to unbalanced bilingualism (e.g., Yip & Mathews, 2007) could result in CLI. Specifically, they predict CLI to occur “when the child has reached a more advanced level of syntactic complexity in one language than in the other” (Paradis & Genesee, 1996, p.3). An example for this is provided in Yip and Mathews’ (2007) longitudinal study on the acquisition of relative clauses in 3 bilingual Cantonese-English children who are Cantonese-dominant. In Cantonese, pronominal relatives emerged slightly earlier than in English, while object relatives emerged simultaneously or earlier than subject relatives. This means that bilingual Cantonese-English children underwent a period in which relative clauses were present in Cantonese but not yet in English. Their results also showed that the same order of emergence was transferred to English “despite it being typologically a dispreferred option¹⁸” (Yip & Mathews, 2007, p. 297). On the other hand, some studies suggest that the influence of the dominant language does not necessarily lead to transfer from the dominant into the weaker language (e.g., Pérez-Leroux et al., 2009). Moreover, van Dijk et al.’s (2022) meta-study revealed that it is rather the strength of CLI, which is determined by language dominance and not its directionality.

Child L2-acquisition transfer models are often based on models put forward for adult L2 acquisition and are usually phrased in structural terms, such as “the early availability of functional categories (e.g., Tns, Agr, Comp) [including higher functional projections, e.g., DP, IP, and CP] and features” (Chilla et al., 2021, p. 230), e.g., the Feature-Assembly Hypothesis¹⁹ (Lardiere, 2008, 2009), and the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou, 2007). Accordingly, CLI effects are to be particularly expected at the morphological spell-out level (Prévost & White, 2000) since features of functional categories, especially those that concern interface phenomena and uninterpretable features (i.e., grammatical features lacking semantic content like case,

¹⁷ Developmental asynchrony is closely linked to the *Bilingual Bootstrapping Hypothesis* (Gawlitsek-Maiwald & Tracy, 1996).

¹⁸ To account for this unexpected order of emergence (i.e., object relatives before subject relatives), Yip and Mathews (2007) argue in line with Diessel and Tomasello (2000, 2005) that those early emerging relative clauses are isolated noun phrases with the linear order of main clauses; hence, their “emergence is facilitated by the isomorphism between object relatives and main clauses ... resulting in a configuration that is conducive to early production” (p. 297).

¹⁹ The Feature-Assembly Hypothesis assumes that successful L2 acquisition depends on reassembly of L2 features that are already available in the L1 into “new functional categories and lexical items”. Hence, convergence is determined by “whether L1 features have the same morpholexical expressions in the L2” and by the L2 learners’ ability to efficiently reconfigure them (Domínguez et al., 2011, p. 2).

grammatical gender, or number agreement on the verb), may be assembled in a non-target-like manner. Moreover, CLI effects are also expected for “features that influence word order properties such as verb placement” (Chilla et al., 2021, p. 231).

In the domain of syntax, transfer of L1 properties has been reported for (early) acquired syntactic phenomena like word order and the placement of arguments and phrasal adjuncts (i.e., modifiers) in relation to the verb. According to Haznedar, “not only lexical categories but also functional categories might transfer into the initial representation of the L2 grammar” (p. 24). For example, in a longitudinal case study with a Turkish-English bilingual child (Erdem) who started acquiring English at the age of 4;1, Haznedar (1995, 1997a, 1997b, 2001) documented transfer effects of Turkish verb placement properties on the development of VP in Erdem’s very early English. A principal syntactic difference between Erdem’s L1 and L2 is that Turkish is a head-final language with a surface OV order in the VP, whereas English is head-initial with surface VO order (Haznedar, 2001). This has consequences for the placement of objects, adverbials, and negation in relation to the verb, e.g., the object in Turkish is preverbal, while negation markers appear in post-verbal positions as in example (1) (Haznedar, 1995, 5 (7)).

- (1) (Ben) *kitap* al -ma -yacağ -ım
 (I) *book* buy -neg -future -1sg
 ‘I will not buy a book’

Erdem’s first 7-8 recordings stretching over three months were dominated by utterances, in which verbs were constantly placed in sentence-final position reflecting an OV setting, e.g., ‘I *something* eating’, ‘would you like to *outside* ball playing?’. Such a pattern is not attested in typical L1 English developmental data. Starting at about recording 8 (after about 5 months of exposure to English), an abrupt change to surface VO order was observed in Erdem’s English utterances, accompanied by a rapid increase of overt subjects which were realized in 94% of obligatory contexts. This led Haznedar to conclude that a resetting of the headedness parameter must have taken place, resulting in a switch of VP headedness from Turkish head-final (SOV) to English head-initial (SVO).

For L2 French, bilingual children are reported to encounter difficulties with the placement of clitics and the formation of questions (Hamann & Belletti, 2006; Prévost et

al., 2014). For L2 German, basic OV word order and SVA were found to be susceptible to CLI (Haberzettl, 2003; Wegener, 1994; see also Döpke, 1998). In the area of morphology, child L2 learners with non-inflecting first languages have been demonstrated to acquire L2 inflectional morphology at a slower pace than children with inflectional languages. For instance, in a longitudinal study stretching over the period of two years, Blom et al. (2012) analyzed language samples of a subset of the children featured in Goldberg et al.'s (2008) data corpus for the use of 3rd person singular inflectional morpheme {-s} (3SG-s) in obligatory contexts. The children were divided into two typological groups depending on whether their L1 inflects for tense and agreement: non-isolating/inflecting L1 (Romanian, Spanish) and isolating L1 (Cantonese, Mandarin). All participating children were born in Canada; however, their systematic exposure to English did not start until they entered preschool. Their results showed that children with inflectionally rich L1s attained 90% accuracy rates on 3SG-s after fewer months of exposure to English than children from isolating languages, who consistently performed worse than their LoE-matched peers from inflecting languages, see also Paradis et al. (2016). Similar findings were reported in Paradis et al.'s (2011) cross-sectional study, not only for 3rd person singular {-s} but also for auxiliary verbs BE and DO suggesting that children with L1s that overtly mark tense/agreement, i.e., finiteness, had an advantage over their peers with isolating L1s. Moreover, the children in Blom et al.'s study appeared to be sensitive to morpho-phonological properties conditioning allomorphy in that they encountered more difficulties marking agreement on stems ending with sibilant sounds. The latter segments are characterized by their lower frequency compared to other types of segments in Dutch (see also Blom & Paradis, 2013; and Rispens & de Bree, 2015 for similar results in Dutch). In the same vein, using controlled elicited production, Blom and Baayen (2013) tested the accuracy of bilingual children (ages 4-9 years) with isolating (Cantonese, Mandarin) and inflectionally rich languages (Moroccan-Arabic/Tarifit, Turkish) on subject-verb-agreement in L2 Dutch in main clauses without (SVO) and with inversion (AdvVS) as well as in relative clauses (SOV). Their results showed that among child L2 learners with lower proficiency²⁰ (intermediate stage of proficiency), those with isolating L1s were less accurate in SVA

²⁰ Effects of CLI are assumed to be strongest at intermediate levels of L2 proficiency, i.e., “when the developmental readiness for transfer” is given, but the L2 learners are yet not proficient enough to perform at ceiling (Blom & Baayen, 2013, p. 804; see also Wode, 1983 and Zobl, 1980). This implies that the influence of the L1 is likely to diminish as cL2 children become more advanced in their L2.

than those with inflecting L1s. In particular, children with isolating L1s struggled with inflectional markings of 2SG and 3SG in main clauses without inversion and with 3SG markings in contexts that involved inversion, where they produced uninflected forms (i.e., omission of finiteness markings). The authors interpret the superior performance of L2 learners from non-isolating languages within the Full Transfer/Full Access hypothesis (FT/FA)²¹ (Schwartz & Sprouse, 1996; White, 1991, 2003), which was originally postulated to account for adult L2 acquisition. Blom and Baayen (2013) argue that although the FT/FA “is phrased in structural terms (e.g., parameter settings)”, the influence of L1 can also be linked to perception or attention (p. 803). For example, child L2 learners with a *pro-drop* L1 (e.g., Arabic and Bulgarian), in which the sentential subject can be identified through inflectional cues on the verb, might generally be more sensitive to verbal inflection markings and distributional cues. Hence, this might allow them to pick up on inflectional cues in the L2 at a faster rate than their peers with isolating L1s. Additionally, finite verbal forms of the type stem+ {-t} were more error-prone than stem+ {-en} forms in the isolating group. Note that stem forms + {-t} would result in a coda cluster in word-final positions, whereas adding the inflectional morpheme {-en} would not. According to the authors, this could have arisen due to the transfer of L1 phonotactic constraints that prohibit the occurrence of consonant clusters in word-final positions in languages like Cantonese and Mandarin (cf. Flege et al., 1995). It should be noted that CLI influence is not limited to the area of morphosyntax as it frequently occurs in other domains like phonetics/phonology (see Kehoe et al., 2004 for results on the acquisition of voicing contrast by German-Spanish bilingual toddlers; Lleó, 2018 on the acquisition of spirantization in Spanish by simultaneous school-age Spanish-German bilingual children; Werker & Tees, 1983 for results on the loss of perception abilities of non-native speech sounds in young children), and semantics (see Ionin et al., 2009 on the acquisition of the distinction between definiteness and specificity).

Findings like the above highlight the importance of considering the child’s L1 when assessing L2 abilities. However, although CLI apparently plays a role in child L2 development, Paradis (2010b) stresses the importance of considering CLI among other variables that contribute to individual variation in bilingual language acquisition, such as

²¹ The FT/FA hypothesis stipulates that the entire L1 grammar, including parameter setting of the L1, represents the initial state of L2 acquisition (Full Transfer), but that UG remains fully accessible for L2 learners throughout the L2 acquisition process (Full Access), which would allow for parameter resetting to the L2 value.

AoO and LoE. For example, CLI effects on the acquisition of article semantics appear to interact with AoO to the L2 as bilingual children fluctuate to a considerably lesser extent than adult L2 learners and late child L2 learners (Chondrogianni & Marinis, 2016; Schwartz & Rovner, 2015). Therefore, Paradis calls for caution as there is a risk for overinterpreting the role of CLI, especially since not all studies found an effect of L1 transfer on the development of L2 (morphology: e.g., Belletti & Hamann, 2004; Paradis et al., 2008; Zdorenko & Paradis, 2008), see also Meisel (2009, 2013, 2018) for arguments against qualitative transfer. Furthermore, CLI appears to be most evident at the early stages of L2 acquisition (e.g., Unsworth, 2005; Zdorenko & Paradis, 2008; see also Schwartz & Sprouse, 1996 and their Full Transfer, Full Access model; see White, 2020, for an overview of such models). Hence, “inhibitory L1 influence” cannot credibly account for the “protracted acquisition of a structure” in bilingual children with more than three years of exposure to the L2 (p. 125).

Moreover, in the case of simultaneous and early successive bilinguals, L2 acquisition begins at a time when L1 still has not reached a stable state grammar, and late-acquired structures have not been mastered. Consequently, late-acquired L1 properties cannot be responsible for certain L2 errors. For example, in the study by Zdorenko and Paradis (2012), the prevailing error type in the Arabic and Spanish L1 groups was the overuse of the definite article *the* in indefinite contexts, especially in specific non-definite contexts. Given that both Arabic and Spanish have article systems similar to English, i.e., they distinguish between definite and indefinite contexts, the latter error type is unlikely to result from L1 transfer and is rather developmental in nature since it is also observed in English L1 acquisition. Interestingly, the Arabic L1 group did not perform any worse than the Spanish L1 group, although Arabic only has an overt article marking definiteness, while indefiniteness is expressed by a zero morpheme. If L1 transfer were at play, one would expect the Spanish L1 children to “transfer their knowledge of the indefinite article into English”, and the Arabic-speaking participants to omit the indefinite article in indefinite/specific and indefinite/non-specific contexts. Yet, the rate of indefinite article omissions in both groups was negligible, suggesting that the “specific transfer of the mapping of the semantic feature [-definite to the respective form] did not take place” (p. 57). The authors argue that the latter error type might have been triggered by the relative complexity of the indefinite feature mapping, rendering it inherently more difficult to acquire, even in monolingual acquisition. Assuming that the Arabic and Spanish-speaking participants “did not have a stabilized knowledge of this mapping in their L1, they had

no pre-assembled feature-morpheme mapping to transfer to their L2”, thus resembling their monolingual English-speaking peers at the outset of acquisition.

Quantity of L2 input

A substantial body of research has explored the differential effects of language input on the acquisition of various linguistic phenomena in relation to input quantity. It has been established that the amount of exposure bilingual children receive in each of their languages is linked to their language abilities in the respective languages during childhood starting from toddlerhood (Pearson et al., 1997) all the way through childhood, especially in the early stages of bilingualism (see Armon-Lotem & Meir, 2019; Unsworth, 2016b).

Different methods exist for quantifying bilingual exposure; see Armon-Lotem and Meir (2019) for an overview. The most straightforward quantitative input factor is the length of exposure to the L2 (LoE) (Paradis, 2019), which measures the approximate number of years since the onset of sustained exposure to the L2. Typically, AoO and LoE are highly correlated (Chondrogianni & Marinis, 2011; Paradis, 2011) as LoE is calculated by subtracting AoO from the child’s chronological age. Nevertheless, they do not necessarily show a linear relationship (Bedore et al., 2016). Many studies reported positive associations between LoE to the L2 and the performance of typically developing bilingual children on various language measures. For example, Chondrogianni and Marinis (2011) investigated the extent to which child internal and language environment factors predicted the performance of monolingual English-speaking children (N = 30; age range: 6;1-8;11; mean(SD): 7;4 (0;8)) and Turkish-English early successive bilingual children (N= 43; age range: 6;2–9;8; mean(SD): 7;8 (1;0)) from low SES backgrounds on standardized English language measures assessing vocabulary (BPVSII, Dunn et al., 1997), receptive grammar (TROG-2, Bishop, 2003), tense morphology (TEGI, Rice & Wexler, 2001) and complex syntax (DELV, Seymour et al., 2005). Their results showed that, alongside AoO and self-reported maternal L2 proficiency levels, LoE predicted 50% of the variance in performance on general lexical measures, suggesting that a longer LoE contributes towards vocabulary development. Additionally, LoE accounted for a meaningful proportion of the variance observed in performance on the TROG-2, the DELV subtest assessing comprehension of complex syntax and 3rd person sg. -s marking on the verb (see also Blom et al., 2012; for similar results on 3rd person sg. tense marking). Interestingly, LoE was not correlated with performance on past tense morphology,

implying that different phenomena might be differently affected by length of exposure. In the same vein, Paradis et al. (2017b) found that, among other factors, LoE contributed to predicting the production of complex syntax in spontaneous language samples of English L2 children from various L1 backgrounds; see also Roesch and Chondrogianni (2016). In a more recent study, Hamann et al. (2020) found that late successive bilingual children with a relatively short LoE to L2-German (LoE: 18-24 months) performed significantly worse than their late successive bilingual peers with LoE > 24 months not only on standardized L2 measures assessing vocabulary and morphosyntax but also on a LITMUS sentence repetition task assessing German morphosyntax.

Although LoE is indicative of the amount of exposure to a given language, it represents a “very crude measure of relative exposure” (Armon-Lotem & Meir, 2019, p. 195) that does not capture day-to-day variations in the amount of exposure to each language over the years. Not to mention that it does not consider how linguistic input is distributed among various contexts and interlocutors. For example, two bilingual children with a similar LoE in terms of years of exposure to the L2 might significantly differ in the relative amount of exposure they receive in the L2 depending on the number of hours they spend at kindergarten or school and whether they engage in playdates in the societal language. A widely used quantitative measure that accounts for variations in the amount of exposure in the two languages is the ‘relative amount of exposure’. It is determined by calculating the (average) amount of exposure a child receives in each of her languages inside and outside the home environment on a daily or a weekly basis and provides “an approximation of the exposure to each language relative to the other” in the form of percentages (Armon-Lotem & Meir, 2019, p. 195). The latter information is usually obtained via detailed parental questionnaires, e.g., ALEQ-4 (Paradis, 2011), documenting language input/output in different contexts and with different interlocutors. Such questionnaires allow the calculation of cumulative exposure (adding up past and current exposure) and the amount of early and current exposure separately. Some studies use very fine-grained calculations of relative exposure, e.g., Unsworth (2013b); see Kaščelan et al. (2022) for an overview of parental questionnaires. A further method for quantifying input quantity is the so-called ‘absolute amount of exposure’, which measures the frequency in which a certain construction occurs in child-directed speech in a given language without considering the other. It differs from relative exposure in that it also captures more qualitative aspects of language input by providing frequency counts of certain phonological, lexical, or grammatical constructions (e.g., De Houwer, 2014). This type

of measure is frequently employed in usage-based and constructivist studies to investigate frequency effects on the acquisition of target language properties (e.g., Ambridge et al., 2015).

Variations in input quantity as estimated by ‘relative amount of exposure’ have been shown to positively influence performance on measures of L2 phonology (e.g., Ruiz-Felter et al., 2016), vocabulary (e.g., Bohman et al., 2010; Cobo-Lewis et al., 2002; Scheele et al., 2010; Thordardottir, 2011), certain areas of morphosyntax (e.g., Blom et al., 2012; Hoff et al., 2012; Gathercole 2002a, 2002b, 2002c; Gathercole & Thomas, 2009; Paradis et al., 2014; Thordardottir, 2015b; Unsworth, 2014), and complex syntax (e.g., Roesch & Chondrogianni, 2016). Gathercole (2002c) and Gathercole et al. (2013) note that input quantity appears to have stronger effects in early years and that by the time children reach middle school (at around age 10;0), performance differences between children with varying amounts of exposure would have largely resolved suggesting that children might have reached “a critical mass of input with respect to relevant linguistic features”, see Beauchamp and MacLeod (2022) for results on school-aged children. A frequent observation is that younger bilingual children with less exposure to the L2 perform significantly worse than their monolingual and bilingual peers with more exposure to the target language. However, it is important to note that the effects of amount of exposure are not across the board and might differ within and across language domains. Especially at younger ages, the impact of amount of exposure may differ at various points within the acquisition process (Gathercole, 2002b; Zurer-Pearson & Amaral, 2014). It has been argued that a certain threshold²² of language input has to be reached for bilingual children to perform at a level commensurate with their monolingual peers and that this threshold depends on whether expressive or receptive abilities are being assessed (Brandeker & Thordardottir, 2015; Gathercole, 2007; Gathercole & Thomas, 2009; Hoff & Core, 2013; Thordardottir, 2011). For instance, in a study on lexical development in simultaneous English-French bilinguals in Montreal, Thordardottir (2011) compared bilingual children with varying exposure profiles, i.e., balanced: 40-60% exposure to English, more English (61-94% English exposure), and more French (5-39% English exposure) to monolingual English- or French-speaking children. Her results revealed a strong link between vocabulary scores and same language experience, with a higher

²² Peña et al. (2011) referred to children with less than 20% of exposure to a language (ca. 17 hours per week) in their participant sample as “functionally monolingual”, see also Pearson et al. (1997).

exposure threshold for expressive vocabulary than for receptive vocabulary. At the group level, around 40% -60% of language exposure was sufficient for performance comparable to monolingual children on receptive vocabulary, whereas 70% exposure was required for monolingual-like performance on expressive vocabulary, see also Cattani et al. (2014) for similar findings. In a further study, Thordardottir (2015b) examined the impact of amount of exposure on the grammatical development of preschool French-English bilingual children aged 3 to 5 years. Using spontaneous language samples in both languages, Thordardottir analyzed mean length of utterance (MLU) in words and morphemes, the accuracy of grammatical morphology, and the diversity of correctly used grammatical morphemes. Interestingly, not all measures featured in the study were equally sensitive to the amount of exposure: MLU was sensitive to bilingual exposure in English but to a lesser extent in French, and the diversity of morphemes emerged as the most sensitive to variations in the amount of exposure. Concerning the weaker correlation between amount of exposure and MLU measures in French, the author argues that this is likely to be linked to “the greater morphological complexity of French compared to English” (Thordardottir, 2015b, p. 110). This implies that the amount of exposure interacts with the linguistic properties being acquired. To examine the latter assumption, Unsworth (2014) investigated input effects on the acquisition of grammatical gender marking on definite determiners and interpretive constraints on different word orders of indefinite objects in Dutch, i.e., scrambling, in a sample of 119 simultaneous English-Dutch bilinguals across seven different age groups (à 14-18 participants per age group, global age range: 5-17 years). The latter two phenomena represent linguistic properties that belong to two distinct domains in Dutch: gender marking (i.e., gender assignment) constitutes a morphosyntactic property with a significant lexical component²³, whereas indefinite object scrambling involves both syntax and (compositional) semantics. Given that scrambling represents a “poverty of the stimulus problem” in that the information crucial for the acquisition of meaning constraints on different word orders is not sufficiently available in the input, input quantity effects were not anticipated for scrambling. In line with previous research (e.g., Bianchi, 2013; Cornips & Hulk, 2008;

²³ According to Pesetsky and Torrego (2007), in languages that display formal gender, nouns are allotted a gender value in the mental lexicon, i.e., they are [+valued] for gender. For example, the noun ‘Tisch/table’ in German bears the value *masculine* for the feature gender, whereas the gender feature for adjectives is lexically unvalued. This means that adjectives get valued as a result of syntactic agreement with the gender feature of the noun.

Stöhr et al., 2012), input effects were expected for gender marking on definite determiners as it is a “lexically specified property of nouns” that has to be acquired on a lexical basis in languages which do not assign gender by means of formal cues, i.e., in languages with a relatively opaque grammatical gender²⁴ like Dutch (Unsworth, 2008). As anticipated, Unsworth (2014) found exposure effects only for gender marking but not for scrambling, suggesting that the acquisition of scrambling does not rely on the amount of exposure but rather “on the availability of the right kind of input” (Armon-Lotem & Meir, 2019, p. 212). The latter findings are consistent with other studies demonstrating that amount of exposure interacts with language-level properties, such as the levels of complexity or opacity²⁵ (e.g., Blom et al., 2012; Nicoladis et al., 2007; Paradis et al., 2014; see also Thomas et al., 2014 on the acquisition of Welsh plural morphology). Moreover, acquisition rates of morphosyntactic constructions might vary depending on the “bilingual child’s dominant language of input” (Paradis et al., 2014, p. 177) as demonstrated in the studies by Paradis and her collaborators (Paradis et al., 2011, Paradis et al., 2014). Hence, the linguistic performance of a bilingual child should always be evaluated in relation to exposure levels to the respective language, considering the linguistic properties of the construction under investigation. Note that some researchers recommend considering ‘input’ and ‘child’s own language output’ separately as output has been shown to have its own contribution to predicting variations in bilingual language performance above and beyond input, e.g., Bedore et al. (2012); Bohman et al. (2010); Goldstein et al. (2010); Paradis (2011) and Ribot et al. (2018). A possible explanation for the special role ‘output’ plays in bilingual language acquisition is that it attracts more input (see Ribot & Hoff, 2014, on code-switching in Spanish-English toddlers). Moreover, Bohman et al. (2010) argue that “using a language (i.e., output) forces the learner to process the language in a way that only hearing it (i.e., input) does not” (p. 339); see Unsworth (2016b) for a discussion. Nevertheless, due to practicality, most

²⁴ In Romance languages like Italian and Spanish, grammatical gender is relatively transparent (i.e., gender assignment relies on the presence of formal cues on the nominal form), whereas in languages like Norwegian and Welsh, it is opaque and tends to be acquired later in monolingual acquisition (Gathercole, 2002b; Kupisch, et al., 2002; Rodina & Westergaard, 2015). Note that transparency of grammatical gender exists on a continuum cross-linguistically and is determined by “(i) the complexity of the system, expressed through syncretism, (ii) the number of gender values, and (iii) regularity of gender agreement” (Velnic, 2020, p. 1; see also Audring, 2014).

²⁵ Opaque constructions are characterized by lower type/token frequency and by inconsistencies in form-to-function mappings. Hence, they require more input and tend to be acquired later than less opaque constructions (Paradis et al., 2014).

questionnaires subsume input and output under ‘language use’ or ‘exposure, e.g., the Questionnaire for parents of bilingual children (PaBiQ, Tuller, 2015).

2.1.2 Bilingual language development across language domains

A frequently asked question by educators and clinical practitioners is how long it takes bilingual children to converge with their monolingual peers. However, setting expectations of bilingual children is complicated by the fact that bilingual children do not converge with monolingual performance in different linguistic sub-domains at the same pace (Oller et al., 2007; Paradis, 2019; Paradis et al., 2011) and tend to show “differential or asynchronous acquisition of the various language domains” (Chondrogianni & Marinis, 2011). For instance, a bilingual child might exhibit average performance in morphosyntax but perform poorly on vocabulary tasks. This is referred to as “profile effects” in bilinguals. Simply put, profile effects “are differences in a child’s performance in distinct areas of the language s/he is learning” (Gathercole, 2014, p.363) that can be observed both across and within language domains (Gathercole, 2002a, 2002b, Oller & Eilers, 2002; Oller et al., 2007).

Profile effects are reported to prevail mainly in less advanced stages of language development. They have been shown to depend on the nature of language aspects/property being acquired (e.g., semantic and morphosyntactic complexity of the forms being acquired) in interaction with child-internal and child-external factors related to the language environment of the child (Bohman et al., 2010; Gathercole, 2007; Oller et al., 2007). According to Gathercole (2017), gaining command of some linguistic aspects requires more exposure than others and takes more time, especially if they are “complex and involve opaque form-function mappings” (p.129). Moreover, domains like the lexicon, in which knowledge is distributed across different contexts (e.g., home vs. school context), and which are highly reliant on piecemeal learning, are more prone to profile effects than knowledge that is not confined to a specific context, like phonology (Bedore et al., 2005; Oller et al., 2007; Peña et al., 2020), where a speedy convergence to target L2 phonology is frequently observed for child bilinguals (cf. McCarthy et al., 2014). For example, in a study on narrative abilities in 21 low-SES bilingual L2 English children (mean age = 58 months) with various L1 backgrounds in Canada, Paradis and Kirova (2014) observed an asynchronous performance profile “in terms of proximity to monolingual norms” for different subskills related to narrative production (p. 342). In

particular, regardless of the amount of exposure, the participating children performed worse on microstructure elements that heavily rely on L2-specific lexical and grammatical knowledge (e.g., utterance length, sentence complexity, and lexical diversity), falling more than 1SD below monolingual norms than on macrostructural elements that are less reliant on language-specific knowledge. The latter include aspects related to story grammar²⁶, such as story structure categories and internal state terms (Gagarina et al., 2016; Gutiérrez-Clellen, 2012; Schneider et al., 2006). The authors argue that story grammar comprises a cognitive linguistic skill that goes beyond sentence level and “could [; thus,] be considered a *shared* as opposed to a *distributed* skill” (p. 343, emphasis in original). Likewise, in Chondrogianni and Marinis’ (2011) study on L2 English development in successive bilingual children with Turkish as L1 from low SES backgrounds, the children displayed differential development profiles across different L2 domains and structures: participants scored lower on receptive vocabulary and (complex) morphosyntax (articles, passive constructions, *wh*-questions) than on tasks assessing general grammar comprehension (TROG-2) and production of tense morphology as measured by the TEGI test. The latter results were also reflected in the proportion of children who were able to converge to monolingual age-appropriate norms in their study: only a third of the participants achieved this for (complex) morphosyntax and vocabulary, whereas half of the children performed within the monolingual norms on tense morphology, and almost two-thirds for global grammar comprehension. Additionally, vocabulary and complex morphosyntax appeared to be more susceptible to the influence of distributed input than tense morphology (see also Oller et al., 2007 for effects of amount of exposure on L2 vocabulary development). In the following sections, I will briefly elaborate on the development of bilingual children’s abilities in vocabulary, complex (morpho)syntax, and phonology, since morphology has already been discussed in relation to factors affecting bilingual language development in the preceding sections.

2.1.2.1 Lexical abilities in typically developing bilingual children

Most evidence for a delay in bilingual language acquisition stems from studies investigating lexical abilities. Ample research has shown that bilingual children lag

²⁶ Essential parts of story grammar (macrostructure elements) include a setting, an initiating event, a goal, an attempt and an outcome, in addition to internal state terms, i.e., “units describing inner thoughts and feelings” (Schneider et al., 2006, p. 226), which serve initiating the goal and expressing the responses (Gagarina et al., 2015).

behind their monolingual age-matched peers on receptive and expressive vocabulary measures, especially when assessed in only one of their languages (e.g., Bialystok et al., 2010b; Goldberg et al., 2008; Hoff et al., 2014; Jackson et al., 2014; Kan & Kohnert, 2005; Marchman et al., 2010; Montanari et al., 2019; Oller et al., 2007; Pearson et al., 1993; Thordardottir & Juliusdottir, 2013). For example, in an aggregate large-scale study featuring 1738 children (772 monolinguals and 966 bilinguals, ages: 3 to 10 years) from different studies using the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), Bialystok et al., (2010b) reported that bilingual children performed significantly worse than their monolingual counterparts across a wide age range with no decrease in the monolingual-bilingual gap with age. Moreover, when performance on different word types was considered for a sub-group of 6-year-olds ($n = 161$), a significant difference emerged between monolinguals and bilinguals for vocabulary related to the home context but not for school-related vocabulary. Bialystok et al. (2010b) attribute the latter results to variations in the bilingual children's amount of exposure to the L2 (see also Scheele et al., 2010). The authors further argue that the observed asymmetry reflects the distributed nature and type of lexical input bilingual children are exposed to outside the school context. While monolinguals and bilinguals are exposed to school-related vocabulary to a similar extent through schooling, home-related vocabulary is mainly offered to bilingual children in their L1, especially in L1-dominant households.

Successive bilingual children approach language acquisition with higher cognitive maturational levels relative to their monolingual and simultaneous bilingual peers, which might allow older L2 children to accumulate vocabulary faster than their younger BiTD peers (Blom & Bosma, 2016; Goldberg et al., 2008). Nevertheless, they have been reported to consistently perform worse than their monolingual peers even after years of extended exposure to the L2 (e.g., Chondrogianni & Marinis, 2011; Goldberg et al., 2008). Moreover, successive child bilinguals have been shown to sometimes underperform simultaneous bilinguals, at least during the first two years of L2 exposure (Hammer et al., 2008), and to perform in the monolingual pathology range when L2 exposure is limited (Paradis, 2010b). Concerning the rate of L2 lexical acquisition in school-age children, results appear to be mixed, as some longitudinal studies report a slower rate for lexical development in L2 children (Jackson et al., 2014), while others report a faster rate (Simos et al., 2014). Stadthagen-González et al. (2013) point out that children with certain language combinations might benefit from the existence of cognates, giving rise to a faster acquisition rate, at least for receptive vocabulary. Thordardottir

(2019) argues that “over the course of the school years, L2 learners are acquiring the L2 at a similar rate as monolinguals” (p. 238). Nevertheless, since vocabulary is acquired on an item-by-item basis and is intimately related to the amount of exposure, as shown by countless studies (e.g., Gathercole et al., 2016; Gathercole & Thomas, 2009; Place & Hoff, 2011; Thordardottir, 2011; see Thordardottir, 2019 for an overview), and given the nature of bilingual children’s “split” input, L2 children constantly face the challenge of trying to catch up to a “moving target” (Cummins, 2000) in that while bilingual children are trying to bridge the vocabulary gap, monolingual children are in the process of expanding their vocabulary. However, once a certain exposure threshold has been reached, especially in conjunction with a dominance shift to the L2, bilingual children are likely to catch up with their monolingual peers, at least on measures of receptive vocabulary, as shown in Hammer et al. (2008) and Thordardottir (2011).

The exceptionally lower performance of bilingual children on vocabulary measures has been ascribed to a combination of factors, including the distributed characteristic of vocabulary knowledge across domains of two languages (Bedore et al., 2005; Peña et al., 2020), reduced input conditions, and low socioeconomic status as a distal factor modifying input quality (see 2.1.1.3). Furthermore, “lexical items are learned independently in each language regardless of whether or not the relevant concept has been acquired”. Hence, unless some overlap exists between the L1 and L2 vocabularies, i.e., “in the form of cognates, bilingual bootstrapping effects between the two languages remain limited” (Unsworth, 2016b, p. 109; see also Singleton, 1999). For instance, as a result of context-specific learning, for some concepts (*signifié* in de Saussure’s terms), bilingual children have translational equivalents, while they can label (*signifiant*) other concepts in only one of their languages (Pearson, 1998). On the other hand, if the bilingual mental lexicon is considered as a whole, vocabulary size in bilingual children is likely to be equivalent to, if not larger than, that of monolingual children (Oller et al., 2007). Thus, assessing vocabulary in only one of the child’s languages with no access to the other can lead to misdiagnosis with language impairment. Even if assessed in both languages separately, vocabulary scores might be very difficult to interpret since the child is very likely to score below monolingual norms if not in the pathology range in each of her languages, especially in the non-dominant one. As a solution, it has been proposed to consider lexical semantic knowledge across both languages of the child (Peña et al., 2016), i.e., using methods that acknowledge the distributed nature of bilingual lexical

knowledge, such as *conceptual scoring*, *total-semantics*²⁷, and the *two-dimensional bilingual coordinate score*, see Peña et al. (2016) for an overview and details on scoring methods. Unlike single-language measures, such methods have been shown to reduce the risk of overidentification of lexical deficits in bilingual children (e.g., Bedore et al., 2005; Core et al., 2013).

2.1.2.2 Development of complex (morpho)syntax in bilingual children

In the previous sections, we have seen that not only vocabulary but also certain aspects of verbal and nominal inflectional morphology represent a locus of difficulty for bilingual children. It has also been shown that some phenomena are more susceptible to effects of variations in the quantity and quality of dual language exposure in interaction with child-internal factors, such as age, AoO, processing abilities, and L1 typology (Unsworth, 2014) with some studies reporting consistent delays over time compared to monolingual acquisition (e.g., Armon-Lotem et al., 2011; Blom et al., 2012; Bohman et al., 2010; Jia & Fuse, 2007; Paradis et al., 2016; Thomas et al., 2014; Unsworth et al., 2014a).

Recently, there has been an increased interest in examining the acquisition of complex sentence constructions by bilingual children. Examples of such complex constructions are relative clauses, *wh*-questions, and topicalized sentences, which are argued to involve aspects of derivational/computational complexity, rendering them more difficult to process compared to less complex structures (Friedmann et al., 2009; Friedmann & Novogrodsky, 2004; Marinis et al., 2007).

The Computational Complexity Hypothesis²⁸ (CCH) was put forward alongside the Derivational Complexity Hypothesis (DCH) (Jakubowicz, 2004, 2005, 2011; Jakubowicz & Tuller, 2008; Jakubowicz & Strik, 2008) to account for the differential difficulty young typically developing children and individuals with atypical language development encounter with computationally complex morphosyntactic constructions across different languages, such as object relative clauses (e.g., Adani, 2011; Adani et al., 2010; Arosio et al., 2011, 2012; Belletti & Contemori, 2010; Delage et al., 2008; Friedmann et al., 2009; Friedmann & Novogrodsky, 2004; Friedmann & Ruigendijk,

²⁷ Total semantics, also known as total vocabulary differs from conceptual scoring in that it also considers translational equivalents, i.e., cross-linguistic synonyms (Core et al., 2013), while conceptual scoring counts overlapping items only once (e.g., Anaya et al., 2018; Thordardottir et al., 2006).

²⁸ Another related hypothesis is the Computational Grammatical Complexity Deficit (CGC; Marshall & van der Lely, 2006; van der Lely, 2004, 2005a).

2017; Hamann & Tuller, 2014), object *wh*-questions (e.g., Biran & Ruigendijk, 2015; De Vincenzi et al., 1999; Jakubowicz & Gutierrez, 2007; van der Lely & Battel, 2003; van der Lely et al., 2011), and accusative object clitics (e.g., Delage et al., 2016; Grüter & Crago, 2012; Mateu, 2015; Tuller et al., 2011). The latter hypotheses arose within Chomsky's (1995, 2000, 2001) Minimalist Program building on the notion of economy²⁹ (see Hulk & Zuckerman, 2000; Jakubowicz & Nash, 2001; Rizzi, 1990, 2000). According to the CCH/DCH, typical and atypical language acquisition is governed by developmental processing constraints, such as working memory capacities that limit the children's ability to handle computationally complex constructions in the initial stages of acquisition. Within the DCH, complexity is defined in terms of internal (i.e., movement³⁰) and external merge operations as specified in the Derivational Complexity Metric³¹ (DCM, Jakubowicz, 2005):

- a. Merging $\alpha_i n$ times gives rise to a less complex derivation than merging $\alpha_i (n+1)$ times.
- b. Internal merge of α gives rise to a less complex derivation than internal merge of $\alpha+\beta$.

In line with Chomsky (2001), the DCM (Jakubowicz, 2011) assumes that sentence derivation proceeds through a series of internal and external merge operations and that uninterpretable feature checking triggers movement, “by means of which the original merging position (where the constituent is interpreted)” and the landing site at which the constituent is pronounced are linked together (Adani, 2012, p. 2). For example, in the case of *wh*-questions, the DCM assumes that the child is sensitive to how often a *wh*-

²⁹ The principle of economy maintains that certain derivations, such as constructions that involve syntactic movement, “are more costly, [i.e., necessitate the mobilization of more cognitive resources], or less economical, than others, and are therefore avoided in the first stage(s) of language acquisition” (Strik, 2009, p. 91).

³⁰ Syntactic movement refers to an operation, by means of which a structure is created through moving “an element from a basic word order (also termed the base-generated order)” (Ruigendijk & Friedmann, 2017, p. 2017).

³¹ Note that the DCM has been subject to criticism by some researchers for issues related to overt vs. covert movement, transparency vs. opacity at the LF (Logical Form) representation (Strik, 2009), and for the distinction between length and structural layers (Soares, 2006). For example, Strik (2009) proposes placing a constraint on the DCH, according to which derivations that are opaque at LF are more complex than derivations transparent at LF, see also Scheidnes and Tuller (2014, p. 279f) for a discussion on the applicability of the DCM to L2 acquisition. Nevertheless, the “basic intuition that this Metric seeks to capture is that an increase in the number of elements manipulated and/or the number of manipulations involved in a derivation leads to increased complexity” (Scheidnes & Tuller, 2014, p. 263).

element (i.e., a constituent) undergoes merge operations and to the number of constituents that merge in order to form a given structure. Hence, to alleviate processing costs, young children, whose WM capacities are still developing³², tend to initially favor less complex options licensed in the target language and incrementally opt for more complex ones as their cognitive abilities progressively mature with age. In other words, with other things being equal, “syntactically less complex derivations are target-consistent (i.e., correctly spelled out at the interfaces) earlier than more complex ones” (Strik, 2009, p. 92), e.g., Delage (2008); Hamann et al. (2007); Hamann and Tuller (2014); Reilly et al. (2004); Rizzi (1990, 2000); Scott and Windsor (2000); Soares (2006). For example, in French, simple matrix *wh*-questions with *wh*-in-situ (e.g., *Tu joues où?* ‘you play where’) are computationally/derivationally less complex than fronted *wh*-questions (e.g., *Où tu joues?* ‘where you play?’), and less complex than *wh*-questions with subject-verb inversion (e.g., *Où joues-tu?* ‘where play you?’). Unlike *wh*-in-situ³³, fronted *wh*-questions involve internal merge, i.e., overt *wh*-movement to SpecCP (i.e., movement of a phrase to the specifier position of the complementizer phrase), whereas those with subject-verb inversion display I-to-C movement (i.e., movement of the finite verb) in addition to overt *wh*-movement (Prévost et al., 2014). Children acquiring French appear to be sensitive to the differential complexity of the latter structures and are reported to favor the *wh*-in-situ strategy over the other two in early stages of acquisition (Hamann, 2000, 2006; Hulk, 1996; Plunkett, 1999). Moreover, object *wh*-questions are argued to be more complex than subject-*wh* questions as they involve movement of the object DP to the very first position of a sentence (i.e., to SpecCP), which is typically reserved for the subject when the linear surface order of constituents is considered. This results in a non-canonical word order, in which the theme precedes the agent rendering thematic role assignment³⁴ more difficult to process than in subject *wh*-questions that follow canonical word order.

³² A similar trajectory is predicted for language-impaired individuals with limited WM abilities, e.g., children with SLI (Jakubowicz, 2011).

³³ In the derivation of *wh*-in-situ questions, there is only one External Merge, whereby neither the *wh*-element nor the verb undergoes syntactic movement (Internal Merge), i.e., “there is no attraction of the *wh*-word and the finite verb to the CP projection[, and] the uninterpretable features in C are checked via Agree” (Prévost et al., 2014, p. 229).

³⁴ Thematic roles are typically assigned according to the base-generated order: the agent role is assigned via the verb to the NP that precedes it, and the theme role is assigned to the NP that follows it. Within Government and Binding theory (Chomsky, 1981), a trace is left behind DPs/NPs that leave their original position indicated by a gap. A thematic role is then assigned by the verb to the trace and subsequently transferred to the constituent that moved via a chain consisting of the moved DP/NP and its trace (Ruigendijk & Friedmann, 2017).

Moreover, the distance between the filler and its gap (trace) is longer in object *wh*-questions than in subject *wh*-questions leading to an additional processing load, as exemplified in (2a & 2b), see Avrutin (2000), Choi and Trueswell (2010), Crain and Fodor (1985), Deevy and Leonard (2004), De Vincenzi et al. (1999), and Frazier and Flores d'Arcais (1989) for processing explanations.

- (2) a. [Who]_i _____iembraces the mother? subject *wh*-question
 b. [Who]_i does the mother embrace _____i? object *wh*-question

A subject-object asymmetry has been reported for the comprehension of *wh*-questions by children across many languages (English: O'Grady, 1997; Tyack & Ingram, 1977; German: Biran & Ruigendijk, 2015; Roesch & Chondrogianni, 2016; Schouwenaars et al., 2018; Dutch: Metz et al., 2010; Schouwenaars et al., 2014; Italian: De Vincenzi et al., 1999; Hebrew: Friedmann et al., 2009; Friedmann & Novogrodsky, 2011; see Sauerland et al., 2016 for a cross-linguistic overview). Similar processing difficulties have also been observed for object topicalization structures. In SVO languages (e.g., English and Hebrew), the agent typically precedes the theme in main clauses. However, in movement-derived structures, as in the case of topicalized sentences, the theme (the syntactic object) is moved across the agent (syntactic subject) to sentence-initial position, i.e., to a higher projection, as in 'The mother_i, the girl greeted _____i', giving rise to non-canonical word order. Such structures are considered to be more complex than canonical sentences that follow the thematic structure 'agent-before-theme' and are reported to be challenging for young typically developing children and populations with atypical language development (Belletti et al., 2012; Biran & Ruigendijk, 2015; Friedmann et al., 2009, 2015; see also Ruigendijk & Friedmann, 2017). In German (genuine V2 language), children cannot rely on the surface word order to assign thematic roles within *wh*-questions and object-topicalized sentences. However, case and subject-verb-agreement (number mismatch) can serve as cues that help resolve the temporary ambiguity in object *wh*-questions and topicalizations, with unambiguous case markings³⁵ being more reliable than SVA (Arosio et al., 2012; Biran & Ruigendijk, 2015; Metz et al., 2010; Roesch & Chondrogianni, 2016; Ruigendijk & Friedmann, 2017; see

³⁵ Note that due to considerable syncretism in the German case paradigm (i.e., when a single form is used to denote several cases) case marking cannot always help disambiguate structures with non-canonical word order in German.

Schouwenaars et al., 2018 for a constraint-based account). For German, Dittmar et al. (2008) found that object topicalizations are comprehended in an adult-like manner only at the age of seven, when children are able to exploit case cues³⁶ to override word order cues when the two cues conflicted (see Arosio et al., 2012 for similar results on object relatives). Likewise, Schouwenaars et al. (2018) reported that not all 7-10 old typically developing German-speaking children in their sample displayed target-like interpretation of object *wh*-questions (see also Stegenwallner-Schütz & Adani, 2017).

As to pronominal clitics, accusative (object) clitic pronouns, in particular 3rd person accusative clitics, have been shown to be acquired later than nominative and reflexive clitics³⁷ in typical L1 acquisition in French (e.g., Delage, 2008; Hamann et al., 1996; Jakubowicz & Rigaut, 2000; Tuller et al., 2011; Zesiger et al., 2010) and some other clitic languages like Greek, Romanian, and Catalan (Avram et al., 2015; Gavarró & Fortón, 2014; Tsimpli & Mastropavlou, 2007). For example, French monolingual TD children are reported to demonstrate an early acquisition phase that completely lacks the 3rd person ACC complement clitics until age 2;6 (Hamann et al., 1996), but interestingly not the homophonous definite determiners '*le/la*' (Jakubowicz et al., 1998). Nominative (subject) clitics, on the other hand, are used productively starting from around age 2 (Hamann, 2003; Hamann et al., 1996; van der Velde et al., 2002; see Coene & Avram, 2011, and Gavarró & Fortón, 2014 for similar findings on Romanian and Catalan, respectively). As a compensational strategy, very young children tend to either use full Determiner Phrases (DPs) or omit the object altogether. For example, van der Velde (2003) found that monolingual French-speaking children aged 3;2 years (mean) omitted accusative clitics at a ratio of 55.9%, replacing them with full lexical DPs in only 20.5% of the cases. Van der Velde further documented a radical drop in the omission rates of

³⁶ Although German L1 children are reported to accurately produce nominative-accusative case distinctions (structural case) relatively early at around the age of 3 (Eisenbeiss et al., 2006, 2009; Klampfer & Korecky-Kröll, 2002; Tracy, 1986), they cannot use this for the construction of the target underlying syntactic structure at this age (Schipke et al., 2012). For instance, using ERP measures, Schipke et al. (2011) revealed that that 3-year-old German L1 children are sensitive to case violations only in the case of double-nominatives, however, when double accusatives are concerned, even 6-year-olds do not show adult-like performance. The ability to exploit case marking for sentence comprehension rather than local semantic cues such as animacy of the referents is argued to start developing after the age of five (Lindner, 2003, Roesch & Chondrogianni, 2015, 2016), while ability to use subject-verb-agreement cues is argued to develop even later (Metz et al., 2010).

³⁷ A number of studies (e.g., Jakubowicz et al., 1998; Jakubowicz & Rigaut, 2001; Zesiger et al., 2010) reported a hierarchy in the order of clitic acquisition, according to which nominative clitics are the first to be acquired followed by reflexive clitics, and accusative clitics, which are acquired last.

ACC clitics as the children grew older (age 4;2: 21.5% omission and 9.3% lexical DPs; age 6;7: 8.3% omission and 7.3% lexical DPs). A decrease in object clitic omission rates with age has also been reported in Hamann et al. (1996), who analyzed longitudinal data of a monolingual French-speaking child ‘Augustin’ (age span: 2;0.2-2;9.3). Once children reach an age at which they try to actively produce French ACC clitics (around age 3 or slightly after)³⁸, they are reported to frequently commit gender errors with a tendency to overapply the default masculine gender in feminine contexts (e.g., Zesiger et al., 2010). In Tuller et al. (2011), incorrect gender marking affected 17% of the ACC clitics produced by 6-year-olds via an elicited production task.

An asymmetry between production and comprehension of ACC clitics has been documented in several studies, with comprehension being better than production (e.g., Grüter, 2006; van der Velde, 2003; Zesiger et al., 2010). For instance, Zesiger et al. (2010) examined the production and comprehension of pronominal and reflexive clitics by 99 typically developing monolingual French-speaking children (age range: 3;5-6;5). In the elicitation task, when syntactically correct³⁹ clitics were considered, the youngest group (4-year-olds) demonstrated high rates of object clitic omissions (21% omission rate), which occurred to a much lesser extent in the 4;6 age group (< 10% omissions) and almost diminished by age 6 (2.5%). Comparable results emerged for subject and reflexive clitics, which were rarely omitted even by the youngest group (less than 10% omission rates in the 4-year-old group). Interestingly, despite not reaching statistical significance, group means for reflexive clitic omission were slightly lower in the 4;6 and 6-year-old groups.

³⁸ Note that depending on the type of data analyzed (spontaneous vs. elicited production/comprehension) and the experimental set up, studies may vary concerning the age reported for of acquisition of accusative clitics. Moreover, ACC clitics avoidance strategies seem to vary according to the source of data: in spontaneous production, low rates of clitic production are accompanied by frequent use lexical DPs (i.e., substitution errors), whereas high rates of ACC clitic omissions prevail in elicited production (e.g., see Pirvulescu, 2006, for a discussion on French data). Jakubowicz et al. (1996) speculate that the high rates of ACC clitic omissions are linked to types of tasks used for data collection: “We believe that the high percentage of object omission [...] is a by-product of the task itself. If these children assumed that French allowed object deletion, they would have omitted objects also in spontaneous interaction data. As shown [...] this did not happen” (p. 384). Nevertheless, there is consensus that accusative clitics emerge later than nominative and reflexive clitics.

³⁹ This type of count does not penalize children when the morphological features ‘gender’ and ‘number’ are not correctly marked on the clitics. Note that this only concerns subject and object clitics since reflexives are “unmarked with respect to those features” (Zesiger et al., 2010, p. 587). This type of count was crucial for enabling a meaningful comparison between subject, object and reflexive clitics, especially that children displayed problems marking the correct gender on both subject and object clitics, which still surfaced until the age of 6 (Zesiger et al., 2010).

By contrast, children showed mastery of comprehension (truth-value-judgment-task) of object clitics with ceiling performance across age groups (> 95.5% correct) on the match condition (where the test sentence corresponded to the pictorial material), and between 87.5% and 93.8%, depending on the age group, on the mismatch conditions (disregarding morphological errors). These results are in sharp contrast to developmental data of children acquiring Germanic languages, such as English and Dutch⁴⁰, who are reported to encounter problems with the interpretation of object pronouns until age 6, also known as the delay of Principle B (Chien & Wexler, 1990). Despite the good performance on comprehension of object clitics in Zesiger et al.'s study, a small yet significant difference emerged in the mismatch conditions across all age groups between the comprehension of reflexive clitics (Binding⁴¹ Principle A), with ceiling performance patterning like subject clitics even in the youngest group, and object clitics (Binding Principle B), with error detection rates ranging between 87.5% in the 4-year-old group to 93.8% in the 6-year-old group.

Aspects of complexity shared by all types of ACC clitics include non-local binding, appearance in a non-canonical position, and co-occurrence with a nominative clitic (see Tuller et al., 2011). A further important factor Zesiger et al. (2010) identified is the type of movement chains involved in the derivation of clitics. As an explanation for the lower performance on object clitics compared to reflexive and subject clitics in their study, Zesiger et al. singled out the chain⁴² configuration arising during the derivation of object⁴³ clitics as the complexity factor rendering them more difficult to process than reflexive clitics. To be precise, the derivation of reflexive clitics “only requires the processing of nested chains⁴⁴” (Zesiger et al., p. 576), whereas the derivation of object

⁴⁰ Note that the morphosyntactic status of pronouns differs between Germanic and Romance languages in that pronouns have a clitic status in Romance languages but not in Germanic languages like English. In Germanic languages, pronouns can receive reference not only through variable binding but also through coreference (Reinhard, 2011; Reuland, 2001). The latter mechanism is assumed to be responsible for the developmental delay reported for Principle B in Germanic languages as it is challenging for children given its pragmatic nature. By contrast, coreference does not constitute an option for the interpretation of pronouns in Romance languages or is only implicated in very restricted contexts. Hence, children acquiring Romance languages are anticipated to show good performance on tasks assessing clitic interpretation (Avrutin & Wexler, 1992; Conroy et al., 2009; Hamann et al., 1996; see Zesiger et al., 2010 for a discussion).

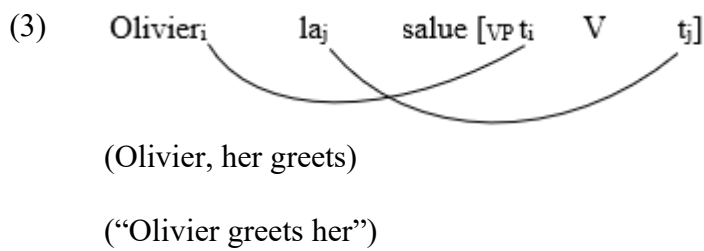
⁴¹ Binding refers to “the links between a pronoun/anaphor and its antecedent” (Zesiger et al., 2010, p. 576).

⁴² “[C]hain relations refer to the links between a moved element and its trace” (Zesiger et al., 2010, p. 576).

⁴³ Another complexity factor that differentiates object and reflexive clitics from subject clitics is their head status, whereas subject clitics are specifiers (Zesiger et al., 2010).

⁴⁴ In reflexives, “the reflexive chain is nested within the subject chain” (Zesiger et al., 2010, p. 597).

clitics requires the processing of crossing chains that arise during the derivation when the “object clitic chain crosses the subject chain⁴⁵” (Zesiger et al., 2010, p. 593), as can be seen in Example (3). In line with other researchers (e.g., Fox & Grodzinsky, 1998), Zesiger et al. ascribe the developmental delay attested for object clitics to the greater computational complexity involved in the derivation of crossed chains compared to nested ones.



Third-person accusative clitics (ACC3) display further characteristics rendering them computationally more complex than other types of ACC clitics, namely grammatical gender marking (in addition to number marking), discourse-participant independence for the reference (hence not marked for animacy) and optionality in specific contexts of French spoken register, i.e., they can be omitted under certain discourse circumstances (Tuller et al., 2011; see Delage et al., 2016, p. 6 for an elaborate discussion of the latter complexity factors). Delage et al. (2016) examined the differential effect of the latter three factors on the production of accusative clitics (early production) by 41 typically developing children (age range: 4-8 years) and found that all three properties influence the production of 3rd person ACC clitics, with the strongest effects for gender. The authors argue that morphological gender marking on the ACC clitics is associated with additional processing costs as “the storage of both the specific gender and the discourse referent of ACC3 may tax working memory resources” in a way that 1st and 2nd person ACC clitics would not (Delage et al., 2016, p. 19). A further interesting finding was that sequences of clitics with gender mismatch appeared to be more problematic than those matched on the gender feature; see Zesiger et al. (2010) for similar findings, but see Avram et al. (2015) on Romanian (a null subject language).

Besides syntactic movement, (*depth of*) *embedding* has been identified as a further factor of computational complexity (Delage et al., 2008; Hamann et al., 2007; see Hamann & Tuller, 2014 for a detailed discussion). Although embedded constructions do

⁴⁵Note that such crossing does not occur in English since the object pronoun remains in post verbal position.

not always involve syntactic movement as in the case of *wh*-questions, passives, or topicalized sentences, they typically involve multiple syntactic operations that vary in number and nature depending on the type of embedded clause, e.g., finite vs. non-finite complement clauses, adjunct and relative clauses (Hamann et al., 2007). In such constructions, the number of External Merge operations interacts with additional complexity factors like the presence of functional material, e.g., verbal inflection and overt complementizers, as well as “tense/mood dependencies between main and embedded clauses”, each of which has its own processing cost (Delage & Frauenfelder, 2020, p. 606). To account for the computational complexity involved in the derivation of embedded clauses, Soares (2006) proposed⁴⁶ adding two further clauses to the DCM (Soares, 2006, as cited in Scheidnes & Tuller, 2014, p. 263):

- c. The computation of a dependent head is more complex than that of a head without this property.
- d. Dependent C: A category C that is specified for tense [Tense] is dependent if and only if C is the domain of the superordinate category that bears a tense feature [Tense].

The general prediction would be that monoclausals, i.e., clauses without embedding, are less complex than embedded clauses, and that the accumulation of syntactic operations results in computationally more complex embedded clauses, which are associated with higher processing costs. For example, adverbial clauses are argued to be computationally less complex than other types of subordinated clauses since they are not selected and not derived by movement; hence, no dependencies arise. Unlike complement and relative clauses, adverbial clauses are “merged to CP or IP (not to V or N) [and hence] involve shallower embedding at phrase edge”. (Hamann et al., 2007, p. 3). On the other hand, finite complement clauses involve dependencies at multiple levels and are thus

⁴⁶ Note that Soares’ (2006) additions to the DCM were based on his research on European Portuguese, specifically in order to predict the order of emergence of left periphery structures.

more complex than adverbial clauses⁴⁷ and non-finite embedded clauses. Moreover, embedding a subordinate clause within another embedded clause (i.e., multiple embedding) is associated with higher processing costs due to deep embedding, e.g., [TP you know [CP that I believe [CP that God exists]]]. The relative complexity of embedded clauses has been shown to correlate with their age of acquisition. Many studies demonstrated that less complex (subordinate) clauses are acquired earlier in L1 acquisition and that the frequency and complexity of embedded sentences (in terms of depth of embedding) increase as the children grow older due to the maturation of WM capacities (e.g., Armon-Lotem, 2005; Delage & Tuller, 2010; Dromi & Berman, 1986; Friedmann & Reznick, 2021; Hamann et al., 2007; Hamann & Tuller, 2014; Reilly et al., 2004; Scott & Windsor, 2000).

Relative clauses (RCs) involve movement and embedding and are thus assumed to be more complex than other types of embedded clauses. Given the recursive nature of relative clauses, a crucial structural complexity factor to consider in RCs is the depth of embedding. This is of particular relevance for some languages like French, which has constructions that appear like relatives but, in reality, are clefts or presentationals, i.e., not genuine relatives, that are acquired relatively early by French L1 children (Delage et al., 2008; Hamann et al., 2007; Hamann & Tuller, 2014). Furthermore, similar to other types of embedded structures, relative clauses that are more deeply embedded are more complex than those that are less deeply embedded (Hamann & Tuller, 2014); see (4).

- (4) a. Tom knows the girl [CP who works in Berlin] Level-1 RC
 b. Tom knows the girl [CP who lives in the city [CP that is the capital of Germany]]

Level-2 RC

⁴⁷ Although adverbial clauses are supposed to be derivationally less complex than complement clauses “in terms of number, nature, and distance of syntactic operations” (Hamann et al., 2007), some studies report that adverbial clauses are acquired later than complement clauses (e.g., Bloom et al., 1989; Dromi & Berman, 1986; Kaplan, 1983; Lustigman & Berman, 2016), with causative adverbial clauses being acquired before conditional and adverbial temporal clauses (e.g., English: Diessel, 2004; Hebrew: Kaplan, 1983; French: Sekali, 2012). As an explanation for the late acquisition of adverbial clauses it has been speculated that “complement clauses ... [constitute] a readily available structure since they are selected by the verb like any other complement”, i.e., they constitute a part of the subcategorization frame of the lexical verb, whereas adverbial clauses, which also involve one embedding layer, are not selected by the lexical verb (Armon-Lotem, 2005, p. 195; see Penner, 1995b for counterarguments from a feature specification point of view). The differential acquisitional rates within adverbial clauses have been attributed to semantic factors, such as the ambiguity of certain subordinating conjunctions (e.g., while: temporal vs. concessive meaning) and syntactic ones like the “degree of [syntactic] integration with the clause they modify” (Endo & Haegeman, 2019; see also Haegeman, 2012).

In the example above, *a.* (Level-1 RC) is less complex than *b.* (Level-2 RC) since it includes an RC in a DP “that is a constituent of the matrix clause”, whereas *b.* contains an RC in a DP “that is a constituent of another subordinate clause”, i.e., more deeply embedded than in a Level-1 RC. This phenomenon is referred to as “relative recursion” (Hamann & Tuller, 2014, p. 49).

A greater difficulty in the comprehension and production of object relatives (ORs) compared to subject relatives (SRs)⁴⁸ has been documented across many languages (e.g., Adani et al., 2010; Arosio et al., 2011, 2012; Biran & Ruigendijk, 2015; Corrêa, 1995; de Villiers et al., 1994; Friedmann & Costa, 2010; Friedmann et al., 2009; Guasti & Cardinaletti, 2003; Rakhlin et al., 2016; Ruigendijk & Friedmann, 2017). For example, Ruigendijk and Friedmann’s (2017) study revealed that monolingual German-speaking children encounter problems with OR comprehension until age eleven, whereas SRs are acquired already at age seven.

Several complexity factors have been invoked to account for the relative difficulty of object-extracted RCs compared to subject-extracted RCs, also referred to as subject-object asymmetry. The first factor is the distance between the constituent that undergoes movement and its trace, i.e., antecedent-gap distance⁴⁹ (Gibson, 1998), which is longer in the case of ORs and more taxing to WM resources since the object has to be held in working memory long enough until one has reached the gap-site, i.e., the position of the trace of the moved constituent, e.g.,

(5) [This is the teacher that the students like ____].



Another complexity factor in ORs is the non-canonical word order that arises due to the movement of the object of the embedded clause to a non-canonical argument position. This is argued to render theta-role-assignment more complex as it has to go through a chain in complex movement configurations as opposed to direct assignment in

⁴⁸ Various domain-general and usage-based accounts have been proposed to explain the subject-object asymmetry in the comprehension of relative clauses, see Levy et al. (2013) for an overview of processing-based accounts, and Brandt et al., 2009 and references therein for a discussion from a frequency/usage-based perspective.

⁴⁹ According to Gibson (1998), “distance is measured by the number of referential expressions intervening between the dependent elements” (Hamann et al., 2007, p. 3).

structures with canonical word order. Moreover, in some languages like French, the possibility of subject-verb inversion, which results in OVS word order within the RC, yields an even more complex structure with two movements (Delage & Frauenfelder, 2019). Concerning the syntactic position of the head of the RC, center-embedded clauses (e.g., the rat that the cat chased ran away) are assumed to be more complex than right-branching RCs and are associated with greater processing demands. The additional processing load arises because the RC head must be maintained in WM while the interrupting RC is being processed until it is subsequently merged with the matrix verb (Gibson, 1998).

Aside from the complexity factors mentioned above, ORs differ from SRs in that they can involve a lexical subject DP, which functions as an intervener between the RC head and its gap site (the moved constituent). The latter property is known as ‘intervention’⁵⁰ and has been identified as the primary source of the difficulty young typically developing children encounter with the comprehension and production of headed object RCs (e.g., Adani, 2011; Belletti & Guasti, 2015; Bentea et al., 2016; Grillo, 2009; Friedmann et al., 2009; Hamann & Tuller, 2015; see Costa et al., 2014 for results on PP object relatives). In their study, Friedman et al. (2009) observed that the nature of the intervening subject and the RC head modulates the degree of complexity of an object relative clause. Specifically, children tended to avoid headed ORs in the cases where the DP of the relative subject and that of the antecedent were both lexically restricted (both contain a lexical NP), giving rise to a configuration as in (6). Typical avoidance strategies in elicited production tasks targeting ORs include the use of subject relatives, which follow the canonical word order or passives (and causatives), especially in older children and adults (e.g., Belletti, 2014; Belletti & Contemori, 2010; Belletti & Rizzi, 2012, 2013; Delage, 2008; Hanke, 2014; see Hamann & Tuller, 2015 for a discussion).

(6) Show me the panda that the mouse is washing <the panda>

+R, +NP +NP +R, +NP

⁵⁰ Intervention can be defined via syntactic features, see Starke (2001) and Rizzi (2004, 2013, 2018), and occurs in special configurations as shown below. If A and B are feature sets, then an intervening constituent with an identical feature set as in (a.) blocks a dependency, whereas a disjunction or a subset relation of the intervening constituent’s features with the features of the target as in (c.) or (b.) allow the dependency (Hamann et al., 2017, p. 11).

a.	* +A..... +A.....<+A>	[problematic or blocked]
b.	ok +A+B..... +A.....<+A+B>	
c.	ok +A..... +B.....<+A>	

According to the Relativized Minimality (RM) account as proposed by Friedmann et al. (2009), when both DPs carry the grammatical feature [+NP], a configuration that violates RM in child grammar arises,⁵¹ in which “the dependency is blocked for children so that they cannot properly comprehend this structure” (Belletti et al., 2012, p. 1054). In example (3), the structural similarity between the DPs (both are lexically restricted DPs) renders the embedded subject DP (*the mouse*) a competitor to the object DP of the matrix clause (*the panda*) in the resolution of the respective structural dependency. Friedmann et al. (2009) propose that children have difficulties computing the subset relations of feature sets⁵² and link this to developmental processing limitations (see also Arosio et al., 2011). Due to their still-developing WM capacities, children are unable to employ adult-like linguistic processes that would allow them to process structures with non-linear dependencies and “have to resort to a number of linear heuristics to process RCs” (Adani, 2012, p. 7). On the other hand, a configuration as in example (6) does not block the derivation of headed ORs in adult grammar⁵³ since the presence of the relative feature [+R] on the gap and RC head creates AB...B... movement configurations that are not problematic for adults, see Rizzi (2013, 2018).

Intervention effects have also been discussed in the context of *wh*-questions to account for the difficulties children and language-impaired individuals encounter with discourse-linked (D-linked, Pesetsky, 1987) or referential object *wh*+NP questions (which-NP questions), as opposed to non-referential bare object *who*-questions (Biran &

⁵¹ In line with Grillo (2008), Adani (2011) argues that RM is similar in child and adult grammars and that intervention effects, i.e., derivation in ORs is blocked due to decay of relevant discourse features in child processing that render the [+R] feature unavailable as a distinctive feature.

⁵² According to Belletti et al. (2012), children are able to compute intersections of feature sets, but subset computation alone is problematic “in configurations in which the relevant feature specification of the intervening subject is a subset of the specification of the moved object. In contrast, children can compute configurations of disjunction of relevant features.” (Friedmann et al., 2017, p. 2).

⁵³ “In adult grammar the DP subject does not interfere in the strict sense of blocking the derivation since the relevant feature shared by the relative operator and its trace is the *wh*-feature of the relative operator, which is not a feature of the subject.” (Hamann et al., 2014, p. 49).

Ruigendijk, 2015⁵⁴; Friedmann et al., 2009, 2017; Friedmann & Novogrodsky, 2011; but see Roesch & Chondrogianni, 2016). Similar to ORs, in object which-NP questions, a lexically restricted subject acts as an intervener between the displaced lexically restricted *Which-NP* object and its trace, as demonstrated in (7).

- (7) [Which boy_{*i*}] did the girl push *t_i*?

Goodluck (2010) criticized the purely syntactic characterization of the intervener as proposed by Friedmann et al. (2009), according to which the mere presence of an ‘NP-restriction’ suffices to trigger intervention effects. As counterarguments for a purely syntactic view, Goodluck mentions semantic-lexical features of the NPs, such as animacy (e.g., Arosio et al., 2011; Brandt et al., 2009; Corrêa, 1995; Grillo, 2008; Kidd et al., 2007; Mak et al., 2002), and a more specific meaning of the NP (e.g., which *lion* vs. which *animal*). Discourse factors (information structure), topichood in particular, have also been shown to reduce the difficulty in the production and comprehension of object-extracted relative clauses (Fox & Thompson, 1990; Mak et al., 2002, 2006, 2008).

Several studies within the Relativized Minimality framework (Rizzi, 1990) examined the role of overt morphological marking on the comprehension of object relatives in morphologically rich languages. Specifically, it has been investigated whether a dissimilarity in the grammatical features (i.e., feature mismatch in gender, number, and case) between the RC head (the moved object) and the embedded subject DP that it crossed (the intervener) facilitates the comprehension of headed ORs (e.g., Adani et al., 2010; Arosio et al., 2012; Belletti et al., 2012; Friedmann et al., 2017). It has been demonstrated that not all types of grammatical feature mismatch facilitate comprehension, i.e., “the system is selective in that not all features are relevant for the computation of similarity” (Friedmann et al., 2017, p. 2). For example, Belletti et al. (2012) compared the effect of morphological gender marking on the comprehension of

⁵⁴ An asymmetry between object-who vs. object-which-NP questions was only attested for the Hebrew data in Biran and Ruigendijk’s (2015) study, whereas object-who questions were as difficult as object which-NP questions for the German participants. The authors attribute this unexpected finding to differences in how object questions are marked in the two languages. In Hebrew both case marking and word order (OSV vs. SVO) serve as cues that help differentiate object from subject questions. However, given the V2 nature of German, object questions follow a non-canonical OVS order “resulting in a mismatch between the syntactic position of the object and their thematic role within the sentence” (Roesch & Chondrogianni, 2016, p. 397). Hence, only case marking helps identify object-questions, especially in cases of unambiguous case marking (see Guasti et al., 2012 for similar results on facilitative effects of unambiguous case marking on OR comprehension in Greek).

headed object relative clauses by 62 monolingual typically developing Hebrew- and Italian-speaking children (age range: 3;9-5;5). Using a sentence-to-picture matching task, Belletti et al. examined whether having a gender mismatch between the RC head (the displaced object) and the intervening embedded subject facilitated the comprehension of ORs in Hebrew and Italian. Their results showed that gender mismatch improved comprehension of RCs in Hebrew but not in Italian. Belletti et al. ascribe the latter finding to the different morphosyntactic status of the gender feature in Hebrew and Italian. In Hebrew, finite verbs are overtly inflected for gender; however, Italian only marks gender on the DP (lexicalized feature) but does not overtly mark gender on the verb. In other words, “gender is part of the featural composition of the inflectional head” in Hebrew (Belletti et al., 2012, p. 1053), i.e., it represents a phi feature that triggers movement of the subject to SpecIP (Shlonsky, 1997). In contrast, verbs inflect for person and number in standard Italian but not for gender, i.e., gender does not belong to the set of inflectional features that attract syntactic movement. Belletti et al. conclude that “only features functioning as attractors for syntactic movement [in a given language] will enter into the computation of syntactic intervention (Belletti et al., 2012, p. 1053). A similar pattern was observed in Biran and Ruigendijk’s (2015) study for the comprehension of object-which questions by young monolingual typically developing Hebrew and German-speaking children (age range: 3;0-6;8), where the gender feature was found to alleviate intervention effects in Hebrew but not in German.

Adani et al. (2010) examined the influence of number and gender mismatch on the comprehension of ORs by typically developing Italian-speaking children aged 5 to 9. Although children performed significantly better in conditions with mismatch-feature-value, the effect size was much larger for the number cue. The authors relate these findings to the different status of number and gender features in Italian; whereas number constitutes “an independent structural head” that drives movement, gender represents a lexicalized feature and hence contributes less to the amelioration of intervention effects (Adani et al., 2014, 2017, p. 816; see also Carminati, 2005; see Metz et al., 2010 for results on Dutch *wh*-questions).

Results on case marking, on the other hand, are subject to debate. While some studies report a facilitative effect for case cues on the comprehension of ORs and object *wh*-questions, specifically when it is unambiguously marked (Arosio et al., 2012; Biran & Ruigendijk, 2015; Guasti et al., 2008; Roesch & Chondrogianni, 2016), others (Bentea, 2012; Bentea et al., 2016; Friedmann et al., 2017) did not find a facilitative effect (see

Friedmann et al., 2017 and Roesch & Chondrogianni, 2016 for discussions) or found a facilitative effect when case was unambiguously marked along number agreement on the verb (Adani et al., 2017; Stegenwallner-Schütz & Adani, 2017). Friedman et al. (2017) state that “only features that are attractors of phrasal movement are used in the computation of intervention, and hence only these features can be used to overcome locality violations” and overriding intervention effects (p. 11). They argue that although case participates in activating the goal and making it available for attraction to the subject position (Chomsky 1995, 2000), as in passive constructions, unaccusatives, and raising structures, it does not constitute an attracting feature for movement and is not a property of the “finite verbal inflection system”. Hence, case cannot partake in the computation of A-bar dependencies (Friedman et al., 2017, P. 11f; see Roesch & Chondrogianni, 2016, for counterarguments). They add that overt case marking on the object can [only] be exploited as a disambiguating cue if “the intervention configuration is properly mastered,” as is the case with unimpaired adults. This implies that young children and language-impaired individuals, who are not yet able to establish an A-bar chain across an intervening full DP subject (i.e., an object dependency chain) in ORs and object which-questions, cannot profit from case marking to resolve the local ambiguity and correctly interpret the structure. Note that all of the latter studies featured young typically developing children (overall age range: 3;6-6;10) whose WM capacities are still developing. According to Hendriks and Koster (2010), young children do not have sufficient processing resources that would enable them to perform complex computations necessary for adult-like performance, especially in structures involving “object dependencies with a lexical restriction” (Bentea & Durrleman, 2014, p. 78). Consequently, they are argued to be unable to efficiently “revise an incorrect initial interpretation” by effectively making use of morphosyntactic disambiguating cues (Schouwenaars et al., 2018, p. 1307f). Hence, Biran and Ruigendijk (2015) point out that investigating older children could reveal a more significant role for case.

As for the relationship between complex morphosyntactic constructions and WM, an interaction with WM and structures featuring multiple complexity aspects can be assumed, especially for young children and individuals with pathologies associated with limited processing abilities, which, as will be seen in the subsequent sections, is of particular relevance for children with SLI (Delage & Frauenfelder, 2020; Jakubowicz, 2011; Montgomery et al., 2010, 2016; Montgomery & Evans, 2009). Such interactions are particularly relevant for feature set computations in constructions with intervention,

especially in the case of object relatives (Arosio et al., 2011; Friedmann et al., 2009; Hamann et al., 2017) and object dependencies with a lexical restriction.

Although the DCH was originally proposed to account for variable acquisition rates of complex constructions in monolingual acquisition, Jakubowicz and Strik (2008) argue that the same principles should apply to bilingual acquisition given that typically developing children are “assumed to have ... an intact language faculty and WM so that operations such as recursion, embedding, and (*wh*-)movement should be intact –given sufficient exposure” to the target language (Hamann et al., 2017, p. 7). Hence, syntactic complexity is presumed to be less susceptible to bilingualism’s effects than other language domains like vocabulary and morphology (Armon-Lotem, 2018; Chiat et al., 2013). Moreover, in the case of sequential bilingual children, transfer of complex constructions from their L1 could also play a facilitative role in the development of L2 complex morphosyntax (e.g., Schwartz, 2004). Support for this comes from studies documenting early emergence and frequent use of complex constructions in typically developing bilingual children (e.g., Belletti & Hamann, 2004; Gavrusseva & Lardiere, 1996; Haznedar, 2003; Scheidnes & Tuller, 2014, 2019). For example, Belletti and Hamann (2004) followed two bilingual children with Italian and German as L1 and French as L2 for 27 months. Both children were exposed to French in nursery schools before age three. Although the German child displayed some transfer phenomena from German in her French spontaneous production, whereas the Italian-speaking child did not, both children produced subordinated constructions “and *wh*-questions (with less complex *wh*-in-situ questions emerging before fronted *wh*-questions as in typical French L1 acquisition) from early on, thus indicating that Tense, Agreement, and Complementizers are available to both” (Belletti & Hamann, 2004, p. 154). Similarly, Paradis et al. (2017b) reported that complex constructions, including subordinated constructions and relative clauses, emerged in less than two years of exposure to L2 English (mean LoE = 17 months) in the spontaneous English productions of ecL2 English learners (mean age = 5;10) investigated in their study. In a further study, Scheidnes and Tuller (2014, 2019) compared the performance of 22 typically developing English-French successive bilinguals with varying AoO of exposure to L2 French (age_{range} at T1: 6;9-12;7, mean (SD): 9;8(1;9); AoO_{range}: 4;0-10;0, mean(SD): 7;0(1;0); LoE_{range} at T1: 0;11-4;0, mean (SD): 2;6(1;1) [years; months]) to a group of 19 age-matched monolingual French-speaking children with SLI (MoSLI) and two monolingual control groups with typical language development (younger and age-matched) at two testing points with an

interval of 12 months (T1 & T2). Analyses of standardized L2 measures (BILO-3C subtests assessing sentence completion, word repetition, and vocabulary, Khomsi et al., 2007) revealed significant overlap between L2 and MoSLI children on subtests assessing vocabulary and sentence completion, especially at T1, with L2 children showing considerably more improvement over time than their MoSLI peers. An interesting finding emerged for analyses of the frequency of subordination (clausal embedding) and depth of embedding in the children's spontaneous speech samples. Despite the high individual variation, the frequency with which L2 children used embedded clauses resembled that of their typically developing monolingual age-matched counterparts. Importantly, they displayed significantly higher embedding rates than their MoSLI peers, where even L2 children with limited exposure to French produced embedded clauses. Furthermore, no correlations were observed with LoE to French. The latter results are in stark contrast to the results on measures of morphosyntactic accuracy (error rates on verbal morphology, object clitics, and grammatical gender marking in spontaneous production), in which L2 children performed similarly to the MoSLI group and significantly worse than their age-matched MoTD peers, showing significant correlations with LoE at T1. Moreover, despite improvement with additional exposure to the L2, error rates remained high for most of the bilingual children after 12 additional months of exposure to the L2. In light of their findings, Scheidnes and Tuller stress that measures of morphosyntactic accuracy and standardized L2 tests are likely to disadvantage bilingual children with limited exposure to the L2 and suggest incorporating measures less sensitive to LoE, like clausal embedding (as an early diagnostic marker), into language measures that aim at disentangling typical from atypical language development in bilingual children.

Similar to monolingual children, bilingual children are shown to be sensitive to the effects of computational complexity. Moreover, complexity effects “may differ according to the properties of the [child's] L1” (Prévost et al., 2014, p. 226). For example, Prévost et al. (2014) reported that Dutch-speaking children acquiring French as L2 were sensitive to computational complexity in that they avoided the structurally more complex questions with *wh*-fronting and subject-verb inversion. However, instead of opting for the derivationally least complex strategy licensed in French, namely questions with *wh*-in-situ, Dutch children mainly produced questions with plain *wh*-fronting (i.e., *wh*-movement without inversion). The authors ascribe this pattern to the possible influence of the “generalized XP-movement to the left periphery” characteristic for Dutch, which is argued to reinforce “optional *wh*-fronting in French” (Prévost et al., 2014, p. 244). This

implies that derivational complexity effects may interact with L1 properties in dual language acquisition.

In order to examine the role of intervention as a complexity factor, Hamann et al. (2017) conducted two studies examining the production of complex constructions in German by monolingual and bilingual children with and without language impairment. The structures in question were long passives, subject relative clauses, and object relative clauses with and without intervention and were elicited using the German LITMUS sentence repetition task (Hamann et al., 2013). An advantage of sentence repetition over other elicitation measures is that it does not easily allow the avoidance of the targeted constructions. The first study featured four typically developing groups (younger MoTDs: 5;6-7;8 yrs., older MoTDs: 8;0-9;10 yrs., younger Russian-German BiTDs: 6;6-7;5 yrs., older Russian-German BiTDs: 7;11-10;5) and a group of monolingual children with SLI (5;7-7;5). Comparing the performance of the five groups revealed that all groups struggled with ORs with intervention, with MoSLI children showing floor performance. Older bilingual children patterned like their monolingual age-matched peers on passives, subject relatives, and object relatives without intervention, but struggled with object relatives with intervention, where performance remained below chance. Frequent errors with ORs with intervention included double case markings, where both the relative pronoun and the embedded DP were marked for accusative or nominative, or resorting to subject relatives. This implies “that WM cannot save a breakdown of the construction: either the relative pronoun is correctly remembered, and the child goes on with the usual order of a subject relative or a subject relative is begun, and then the child remembers the case of the lexical DP” (Hamann et al., 2017, p. 21). Interestingly, the children did not attempt to produce passives instead of object relatives. A similar picture emerged for the younger typically developing groups: MoTDs and BiTDs patterned alike, outperforming the MoSLI group on all structures. However, the younger typically developing children appeared to be more sensitive to the effects of computational complexity than their older peers in that they had slightly higher error rates on ORs without intervention compared to long passives⁵⁵ and subject relatives, which were well mastered. Similar to their older peers, they experienced exceptional difficulties with ORs with intervention. In line with Belletti and Rizzi (2012) and Collins (2005), the authors argue that passives, against their

⁵⁵ Note that Hamann et al. (2017) disregarded a recurrent case error on the PP in long passives, in which the dative case marking on the complement of the prepositional phrase was replaced by accusative case marking.

surface appearance, do not involve intervention and that non-canonical word-order is unlikely to be the primary cause of difficulties with object relatives with an intervener.

In another study on German complex constructions, Öwerdieck et al. (2021) investigated how late successive Arabic-German bilingual children (age range: 8;7-13;1 years; mean age: 9;4; AoO > 5;0) compared to their monolingual German-speaking age-matched peers in production and comprehension of German relative clauses. Production was examined longitudinally with an interval of ≥ 12 months (mean LoE at T1: 23 months; T2: 37 months), whereas comprehension was only assessed at the second measuring point. Production was assessed using an adaptation of the elicited RC production task known as ADIF (Novogrodsky & Friedmann, 2006) and the German LITMUS SRT (including the same sentences featured in Hamann et al., 2017). Comprehension of subject and object RCs was assessed via a picture-selection task (character selection) integrating case (on the relative pronoun vs. on the embedded DP) and number (singular vs. plural) as disambiguating cues. In line with previous studies, a clear subject-object asymmetry was attested at both measuring points in both monolingual and bilingual groups in the elicited production task. Examination of repair strategies revealed that MoTDs predominantly used passives as an avoidance strategy. At T1, BiTDs were found to avoid embedding by mainly using monoclausals or employing the resumptive strategy in their attempts to produce object RCs, which prevails at T2. A clear improvement in the performance of the BiTDs was observed at T2: passives also emerged as an avoidance strategy, and case errors in relatives decreased, implying that “case marking gets stabilized” (Öwerdieck et al., 2021, p. 623). However, this did not result in target-like production of object RCs, most probably due to intervention effects. Other avoidance strategies were also observed, like the use of coordination or *wo*-relatives. As to sentence repetition, in contrast to MoTDs, who only struggled with ORs with intervention, intervention effects did not surface at T1, as BiTDs performed poorly on all types of relatives. At T2, a significant improvement was reported for SRs, whereas the repetition accuracy of both OR types was still low, implying that late L2 children struggled with non-canonical word order. Öwerdieck et al. (2021) argue that intervention effects might have been “obliterated by case difficulties” frequently encountered by bilinguals (p. 623). Öwerdieck et al. point out that despite their lower performance in RC production, examination of avoidance strategies can still indicate typical language development, especially since performance appears to improve with more exposure to the L2. Results of the comprehension task were encouraging as BiTD children showed

mastery of subject relatives, performing almost at ceiling, similarly to their monolingual peers. Concerning object relatives, if group means are considered, the MoTD group (87.5% accuracy score) appears to outperform the BiTD group (73.2% accuracy score); however, the between-group difference does not reach statistical significance. A closer inspection of the bilingual data revealed that almost all BiTDs performed on par with their MoTD peers, except for two children with AoO > 10 years from a very low SES background. Thus, it can be said that RC comprehension offers a promising area for establishing typical language development in older late successive bilingual children.

Hamann et al.'s (2017) second study compared the performance of BiTD (5;8-8;11 yrs.) and BiSLI children (5;6-8;5) with Arabic, European Portuguese, or Turkish as first languages to age-matched MoTD and MoSLI groups not only on the four structures assessed in Study 1 but on all constructions of the German LITMUS SRT. Their results showed similar performance for MoTD and BiTD children who outperformed their MoSLI and BiSLI peers, who also patterned alike. The latter results were replicated in subsequent studies on the German SRT featuring larger samples and controlling for different variables such as language dominance and L1 influence (Abed Ibrahim & Fekete, 2019; Chilla et al., 2021; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018). Since the German LITMUS SRT was created using complex structures such as object *wh*-questions, subject and object relative clauses, complement clauses, passives, and topicalization, one can assume that language development in monolingual and bilingual SLI is fundamentally different from bilingual development.

The findings of the studies presented in this section have shown that aspects of computational complexity do not automatically translate to avoidance or late emergence of complex structures in typical child bilinguals. This confirms “the prediction that typical bilinguals have an intact language faculty and working memory and can deal with movement, embedding, and recursion” (Hamann et al., 2017, p. 28) and shows that computationally complex constructions represent a relative area of strength in typical bilingual language development in comparison to vocabulary and inflectional morphology, which require much more exposure to the L2.

2.1.2.3 Phonological abilities in bilingual children

Given the focus of the studies featured in this thesis, the description of bilingual children's phonological abilities is restricted to segmental and syllable structure phenomena of

direct relevance to this work. For an overview on the acquisition of suprasegmental phonology in bilingual children, see Pronina and Prieto (2024).

With respect to early bilingual segmental phonology, there is converging evidence from perception and production studies confirming “the early separation but interconnected nature of the emerging phonological systems” (Bosch, 2024, p. 350). Especially at younger ages, cross-language interaction effects between the input languages are frequently observed in the form of acceleration, deceleration, and bidirectional transfer (de Almeida et al., 2012; Fabiano-Smith & Goldstein, 2010; Keffala et al., 2018; Paradis, 2001; Paradis & Genesee, 1996; Tamburelli et al., 2015; see Bosch, 2024, for an overview). Several interrelated factors have been proposed to account for interaction effects between emerging phonological systems. These include the phonological properties of the input languages and their typological proximity (e.g., segmental repertoires, relative complexity of vowel and consonant systems, phonotactic constraints, markedness, phonological saliency, syllabic structure, and suprasegmental properties), and language environment factors, such as age of exposure, quantity and quality of L1/L2 exposure and language dominance, particularly in sequential bilinguals (Birdsong & Amengual, 2024; Bosch, 2024). These factors interact with child-level factors, including auditory perception, motor control skills, and their interconnection (perception-production link), as well as general verbal aptitude and cognitive skills (attention to speech, word-form detection, and expressive lexical growth) that support segmental and structural phonological knowledge within and across languages (Keffala et al., 2020; Kehoe & Havy, 2019).

A growing body of literature has identified differences in the developmental trajectories between monolingual and bilingual children, especially during preschool years. However, despite qualitative differences, interindividual variability, and transient delays in some language-specific phonological and phonetic aspects, e.g., marked segments (Fox-Boyer & Salgert, 2014), VOT (Fabiano-Smith & Bunta, 2012; Kupisch et al., 2021), and syllable structure (Keffala et al., 2018), converging cross-linguistic evidence indicates no global segmental learning disadvantage for simultaneous and early-successive bilingual children compared with monolinguals (Bosch, 2024; Bunta & Ingram, 2007; Keffala et al., 2020; see Fox-Boyer & Salgert, 2014; Hambly et al., 2013 for overviews). Below is a brief overview of selected studies on the acquisition of segmental phonology in simultaneous and early-successive bilinguals.

In a pseudo longitudinal study, Albrecht (2017) examined phonological development in 84 simultaneous and early-successive Turkish–German bilingual children (ages 3;0–5;5) in both languages, using the PLAKSS picture-naming task (Fox-Boyer, 2014). An upper limit of 5;5 was chosen because monolingual children typically acquire the full phoneme inventories by that age in both languages (Turkish: 3;11; German: 4;11; Ünsal & Fox, 2002). Albrecht’s results showed that most of the observed phonological processes, particularly syllable structure and consonantal processes, reflected developmental physiological phonological processes that occur in typical monolingual phonological development in German and Turkish, whereas processes atypical for monolingual acquisition, such as lateralization of /ʁ/ (see Fox-Boyer, 2016; Fox-Boyer et al., 2020; Fox & Dodd, 1999) and insertion of an epenthetic schwa in CC clusters were negligible in frequency. Nevertheless, cross-language interaction effects were evident in the form of qualitative differences and a slower acquisition rate for certain German-specific properties, which Albrecht attributes to L1-to-L2 transfer effects. For example, the bilingual children required an additional six to twelve months to overcome the reduction of syllable-initial consonant clusters and fronting of /f/ and /ç/. Similarly, a two-to-three-year delay was observed for the acquisition of syllable-final consonant clusters, along with the deaffrication of /ts/ and /pf/, which are not present in the Turkish phonological inventory. Using a pseudoword repetition task, Grimm and Domahs (2025) compared monolingual German-speaking children’s production of word-initial and word-final consonant clusters to that of early successive bilingual children (ages 4;5 to 5;0) acquiring German as an L2. As possible sources for cross-linguistic influence (CLI) effects from the heritage language, the authors considered the maximal number of consonants permissible in syllable margins (onset and coda), the segmental make-up of clusters, and stress patterns. Their results indicated high accuracy rates and no CLI effects for onset clusters, indicative of a high maturational degree. Conversely, word-final clusters were produced less accurately, and CLI effects were found. The authors ascribe this to the smaller sonority distances between the segments of the final clusters compared to the word-initial ones.

Regarding vocalic phonological processes, vowel substitutions were frequently observed, particularly those involving the front mid vowel /e/, the central mid vowel /ə/, diphthongs, and phonemic vowel contrasts not attested in Turkish (see also Fox-Boyer & Salgert, 2014; Melzer et al., 2018; Yavuz & Balçı, 2011). In the same vein, Kehoe (2002) analyzed spontaneous speech samples from three Spanish–German bilingual toddlers and

found that the bilingual children showed a delay in acquiring the German phonemic vowel length contrast compared with monolingual German peers, while their Spanish vowels were unaffected. Kehoe attributes this pattern to the greater markedness of German vowel length and to reduced or ambiguous input in bilingual settings, leading to cross-language influence or deceleration for the more marked phenomenon. In a further study, Rinker et al. (2010) compared preschool monolingual German-speaking and early-successive Turkish–German bilingual children on the sensitivity to German phonemic vowel contrasts using the oddball paradigm. They found reduced neural sensitivity, indexed by the mismatch negativity (MMN) event-related potential, to German-specific vowel contrasts in bilinguals, but similar sensitivity to vowel contrasts that occur in both the Turkish and German phonological systems (see also Rinker et al., 2017).

Fox-Boyer and Salgert (2014) stress that the input-sensitive delays and language contact effects reported above typically diminish with more exposure, and that realignments can occur during the acquisition process; hence, they do not constitute instances of impaired phonological development. This is confirmed by a growing body of cross-linguistic empirical research showing that bilingual children’s (segmental) phonological abilities improve significantly with age. By the end of the preschool years (around 5–6 years), bilingual children typically have acquired the phonological systems of both languages and are no longer distinguishable from their monolingual peers, at least in the dominant language (Fox-Boyer et al., 2020; Fox-Boyer & Salgert, 2014; Keffala et al., 2020; Ünsal & Fox, 2002; see Bosch, 2024, for an overview). In light of these findings, Fox-Boyer (2016) cautions against pathologizing bilingual phonological patterns, in particular, misinterpreting vowel inconsistencies and segmental acquisition delays as signs of phonological impairment in bilingual children, especially in the preschool years (see also Babatsouli, 2024). Importantly, these differences should be viewed from the perspective of “cross-linguistic variability [as] a result of adaptive processes to the special (richer, more variable) dual language input properties characteristic of bilingual learning contexts” (Bosch, 2024, p. 354).

One important finding in the literature documenting segmental phonological development in simultaneous and early successive bilingual children aged 3 to 6 years is that phonological ability in one language is not only associated with language ability in the same language, especially expressive lexical abilities (e.g., Goldstein et al., 2010; Kehoe & Giradier, 2020), but also with language ability in the other language (i.e., between-language cross-domain interactions). Thus, it differs from vocabulary and

morphosyntax, which predominantly involve within-language cross-domain interactions (e.g., Hoff et al., 2018b). For example, Kehoe and Havy (2019) examined L2 French phonological development in French-speaking bilingual children, testing both French vocabulary and total vocabulary as predictors. Total vocabulary significantly predicted some phonological measures (e.g., percentage consonants correct and word-final consonant accuracy), whereas French vocabulary did not significantly contribute to explaining the variance. In a large-scale study, Keffala et al. (2020) examined predictors of phonological abilities in 695 Spanish–English bilingual children (ages 3:0–6;0) and found that cross-linguistic phonological skills were the strongest predictor of consonant accuracy in each language. Similar associations between the phonological systems of the two languages have been reported by Cooperson et al. (2013), Montanari et al. (2018), and Scarpino et al. (2019). Kehoe et al. (2021) argue that phonology acts differently from vocabulary and aspects of morphosyntax, which are more prone to profile effects (Oller et al., 2007), since they have to be learned on a “language-specific basis” and are highly reliant on the amount of exposure to the respective language (Hoff et al., 2012, 2014). Phonology, on the other hand, appears to be more robust (Hambly et al., 2013) and less susceptible to reduced input effects. A possible explanation for this is that bilingual phonological ability may be underpinned by a shared, language-general articulatory subsystem alongside language-specific phonological components, which “resemble language-general-like components” due to the substantial cross-linguistic overlap between phonetic inventories and phonotactic patterns (Kehoe et al., 2021). Thus, even with reduced exposure to one language, a child can compensate by bootstrapping from the phonology of the other language. Also, other language domains associated with better phonological skills within one language can have a positive impact on the development of phonological skills in the child’s other language.

Given the robust relations between phonology and cross-language general language skills and its relative resilience to reduced split-input effects compared with domains such as vocabulary and inflectional morphology, segmental phonology could be exploited for diagnostic purposes in simultaneous and early successive bilingual children. This is especially relevant for children approaching school age and beyond, covering the age range investigated in this thesis.

2.1.3 Bilingualism and executive function

Dual language acquisition has often been linked to enhanced cognitive processing in bilingual individuals, in particular, executive functions, and is further argued to be a protective factor against cognitive decline associated with aging (Bialystok, 2009; Bialystok et al., 2007, 2012; Blom et al., 2014; Carlson & Metzoff, 2008). The bilingual experience is said to alter “language representations and brain structure that underlie language acquisition or processing” (Bialystok, 2017, p. 3). This line of argumentation is supported by research showing bilingual children to have greater cognitive flexibility relative to monolingual age-peers (see Nicoladis, 2016), which brings about linguistic consequences. For instance, compared to monolingual peers, bilingual children are reported to be more willing to accept that a lexeme can be polysemic (Bialystok et al., 2010a; Davidson et al., 1997) and to be less inclined to adhere to the principle of mutual exclusivity (Markman & Wachtel, 1988) when mapping new words to unknown objects (Bialystok et al., 2010a; Davidson & Tell, 2005). According to Byers-Heinlein and Werker (2009, 2013), bilingual children start exhibiting this type of flexibility once translation equivalents emerge in their vocabularies (see also Bialystok et al., 2009, p. 91 for a discussion). As a result, bilingual children show higher flexibility in interpreting novel words when compared to their monolingual peers (Marinova-Todd, 2012). For example, a study by Yoshida et al. (2011) using a novel word learning task revealed that 3-year-old bilingual children performed better than monolinguals in mapping novel adjectives to a property, a task known to be cross-linguistically challenging for young children and is argued to be linked to young children’s prepotency for assigning novel words to the category of nouns rather than to a different word-class (cf. Gasser & Smith, 1998). Yoshida et al. attribute their findings to enhanced attentional control in bilinguals, which better predicted bilingual performance of their participants than conceptual vocabulary size (see also Kapa & Colombo, 2014).

A frequently asked question is whether bilingualism confers nonverbal cognitive benefits beyond linguistic ones, i.e., whether bilingualism enhances generic executive functions necessary for performing non-linguistic tasks (see, e.g., Special Issue in the *Journal of Cultural and Cognitive Sciences*, Mishra & Abutalebi, 2020). A bilingual advantage over monolinguals has been documented across a broad range of tasks assessing generic executive functions, especially tasks requiring “controlled or selective attention to avoid interference” (Bialystok, 2018, p. 283), i.e., conflict processing, e.g.,

Simon or Stroop tasks (Bialystok, 1999; Bialystok & Viswanathan, 2009; Engel de Abreu et al., 2012; see Lowe et al., 2021, p. 1119 for a detailed overview of various EF-tasks). The cognitive benefits of bilingualism are believed to “emerge from enhanced domain-general attention to jointly-activated competing systems” during language use (Bialystok, 2017, p. 37; see Kroll et al., 2015 for a review). In order to switch between their two (or more) languages depending on the interactional context while avoiding interference from the non-target language during target-language production, bilinguals constantly need to control and monitor one language while using the other (Costa et al., 2008; Kroll et al., 2012; Poarch & van Hell, 2012). Moreover, not only do bilinguals have to avoid interference from the language not in use but also from the target language (Declerck & Philipp, 2015), e.g., at the lexical (Kroll et al., 2008) and the syntactic level (Hatzidaki et al., 2011; Kroll et al., 2012). This is argued to be regulated by a “rapid and efficient selection mechanism” (Bialystok, 2018, p.284), which is assumed to be “part of the domain-general executive function system” (Bialystok, 2018, p.284) consisting of a set of higher-level general-purpose cognitive processes (thinking and reasoning) known as executive functions (EFs). EFs are responsible for managing information and regulating thought and (goal-directed) behavior. The latter include shifting (also called cognitive flexibility), inhibition, planning, updating (working memory), and selective attention, and are assumed to be located in the frontal lobe (supported by subcortical regions) (Bialystok, 2018; Friedman et al., 2006; Henry et al., 2012; Miyake et al., 2000; Miyake & Friedman, 2012; Zelazo et al., 2003). Of particular relevance are shifting, inhibition (inhibitory control), selective attention, and updating (operationalized as WM) since these processes relate to bilingual language selection and are hence expected to be impacted by bilingualism (see Bialystok, 2017).

Shifting concerns the ability to navigate fluently between different mental sets or tasks. This also includes shifting perspectives and switching attention from one thing to another (Best & Miller, 2010; Diamond, 2013; Henry, 2012; Miyake et al., 2000; Miyake & Friedman, 2012). Other known labels for shifting include (task-)switching (Henry, 2012) and attentional or cognitive flexibility (Cañas et al., 2003; see also Blom et al., 2021a). Nonverbal shifting is commonly evaluated via set-shifting tasks (Dajani & Uddin, 2015) that “assess the ability to change previously learned behaviors if these are no longer relevant” (Blom et al., 2021a, p.2). Examples of such tasks are object or dimensional change card sorting tasks, in which cards must be sorted according to different criteria (dimensions) such as shape, number, or color, where the relevant feature varies across the

trials (Iluz-Cohen & Armon-Lotem, 2013; Smidts et al., 2004). It should be noted that tasks assessing shifting “also require minimum ability to inhibit previously activated mental sets in order to perform (switch to a new operation)” (de López & Baker, 2015, p. 279; see also De Cat et al., 2018, pp. 21-22). Bialystok (2018) points out that in addition to inhibition, selective attention and updating are also involved in dimensional change card sorting tasks since the child needs to “selectively attend to the relevant dimension, inhibit attention to the other feature, hold a sorting rule in mind and update it for the post-switch phase” (p.291). Another method used for evaluating shifting is the nonexistent object drawing task (see Blom et al., 2021a, for details; Karmiloff-Smith, 1990). During the task, children can modify size, shape, and location, and delete or add elements. Several studies have reported better performance on shifting tasks in child bilinguals than monolinguals (e.g., Adi-Japha et al., 2010; Bialystok & Martin, 2004; Bialystok & Viswanathan, 2009; Carlson & Meltzoff, 2008; Kalashnikova & Mattock, 2014; Okanda et al., 2010). Interestingly, the reported advantages do not appear to be linked to SES or cultural background and are found for various language combinations (e.g., English-Spanish: Barac & Bialystok, 2012; Carlson & Meltzoff, 2008; English-Tamil: Bialystok & Viswanathan, 2009). On the other hand, some studies did not find a bilingual advantage for shifting (e.g., Gathercole et al., 2014).

Updating is operationalized as working memory⁵⁶ (WM) and refers to the ability to retain task-relevant information in a temporarily highly activated and accessible state while performing cognitive operations (cognitive processing), i.e., it involves simultaneous storage as well as manipulation and processing of the stored information (Baddeley, 2000; Cowan et al., 2005, 2014; Gathercole et al., 2004a; Hambrick et al., 2005; Montgomery et al., 2010). WM plays an essential role in a variety of higher-order cognitive tasks, such as fluid intelligence (Engel de Abreu et al., 2010) and mathematical abilities (Bull et al., 2008; Geary et al., 2004; Swanson, 2006) and is associated with

⁵⁶ Working memory represents one of the two subsystems, of which *memory* consists. The other subsystem is long-term memory (LTM) and “reflects an individual’s permanent knowledge base” (Montgomery et al., 2018a, p. 2951) and a close relationship is argued to exist between WM and LTM, especially in typically developing populations (Baddeley, 2012; Baddeley et al., 2020; Cowan et al., 2012; Loaiza & Camos, 2018). “Language (crystallized) knowledge resides in LTM and includes phonological, morphological, lexical and syntactic representations” (Montgomery et al., 2018a, p. 2951). Lexical-semantic knowledge is said to reside in the declarative LTM, while rule-governed syntactic knowledge comprises part of the procedural LTM (Ullman, 2001, 2004).

activity in the prefrontal cortex (Bunge et al., 2005; see also Chung et al., 2013 for an overview of neuroimaging studies).

Despite the existence of different theoretical frameworks for conceptualizing WM (see Courage & Cowan, 2022; Montgomery et al., 2019 for overviews), such as Cowan's Embedded Processes model (Cowan et al., 2005, 2012, 2014), the dual-storage model (Engle et al., 1999; Kane et al., 2004) and the time-based resource-sharing model (Barrouillet et al., 2008), a substantial amount of psycholinguistic research on the development of WM in children has been carried out within Baddeley's multicomponent WM model (Baddeley, 2012; Baddeley & Hitch, 1974; Baddeley & Logie, 1999). The most recent version of the WM multicomponent model (Baddeley, 2012, Baddeley et al., 2020) distinguishes four functionally distinct yet interactive subcomponents (Baddeley, 2000): (a) the phonological loop and (b) the visuospatial sketchpad, which represent two separate domain-specific memory subsystems (short memory stores), (c) a domain-general central executive⁵⁷ and (d) the episodic buffer.

The phonological loop is responsible for the temporary retention and processing of acoustic/phonological and verbal information and thus corresponds to verbal short-term memory (VSTM), also referred to as phonological short-term memory (pSTM). In addition to a phonological store, which can maintain phonological material for a few seconds until the phonological trace starts to decay, the phonological loop includes a subvocal rehearsal mechanism that allows maintaining phonological representations and rehearsing phonological material in order to refresh the memory trace. The latter mechanism can also be used for recoding "visually presented information allowing access to the loop" (Archibald & Harder Griebeling, 2016, p. 253; Baddeley, 2003, 2012). The capacity of the phonological loop is typically measured via simple serial recall tasks, i.e., simple-span tasks such as forward digit span and nonword repetition (Barrouillet & Camos, 2007; Coady & Evans, 2008; Gathercole, 2006b; Richardson, 2007). Such tasks are claimed to only involve temporary storage and immediate retrieval of verbal information and thus tap verbal short-term memory (e.g., Coady & Evans, 2008) or phonological working memory (PWM) as referred to by some scholars (e.g., Pierce et al., 2017a, 2017b). Performance on simple spans has been shown to incrementally increase with age, particularly between two and nine years of age, and continues to develop until

⁵⁷ Despite its involvement in the "attentional selection of perceptual information", the "central executive" is principally concerned with internally oriented attentional control processes" (Baddeley et al., 2020, p. 11).

adolescence, with crucial changes occurring in the rehearsal mechanism between ages six and eight (Barrouillet & Camos, 2007; Tam et al., 2010). On the other hand, the visuospatial sketchpad is responsible for storing visuospatial information. It is typically measured using tasks with visual patterns or figures requiring the immediate recall of spatial positions and paths (cf. Logie, 1995). An example of such tasks is the Mr. Peanut short-term spatial memory task (adapted from Case, 1985; see de Ribaupierre & Bailleux, 1994 for details).

A further WM component is the central executive, which is a limited-capacity attentional control mechanism (controlled attention) responsible for coordinating and controlling activities of the latter so-called “slave” subsystems within working and episodic memory (Baddeley, 2007; see also Pierce et al., 2017a). Controlled attention involves abilities like “allocating attention resources to different components of WM or other cognitive systems through sustaining focal attention and dividing or switching attention” (Montgomery et al., 2019, p. 3809). The capacity of the central executive, i.e., WM capacity, is assessed using tasks that involve concurrent retention and processing of information briefly stored in one of the two domain-specific storage components (verbal/phonological STM or visuospatial STM) by the domain-general central executive. An important constraint on WM performance is the “ability to shift attentional energy in order to meet both the processing demands of working memory and to refresh or rehearse the material in storage” (Archibald & Harder Griebeling, 2016, p. 253). Popular psychometric measures of verbal WM (VWM) include complex memory span tasks (see Conway et al., 2005 for an overview), e.g., backward digit span, counting span, reading and listening span tasks (Barrouillet & Camos, 2007; Daneman & Carpenter, 1980; Gathercole et al., 2004b; Gaulin & Campbell, 1994), all of which pose significant processing demands in addition to verbal storage capacity (Gathercole et al., 2004a). Compared to simple span tasks, performance on complex span tasks has been reported to show a relatively linear improvement with age and exhibit rapid initial improvements until it levels off by the age of nine (Barrouillet & Camos, 2007) and reaches a stable plateau between ages 14 and 15 (Gathercole et al., 2004a). Nonverbal WM capacity is usually measured with visuospatial complex memory tasks, i.e., spatial span tasks like the Odd One Out task (Henry, 2001; for details, see Henry et al., 2012, p. 40). For example, in the Odd One Out task, the child is presented with three pictorial elements (shapes) embedded in a sequence of boxes and is requested to identify the odd one out, memorize its position and then identify its exact location on a picture with empty boxes. As with

verbal WM tasks, complex visuospatial WM tasks require both storage and processing of the visuospatial information, where the cognitive load is manipulated throughout the trials by alternating storage and processing demands. Hence, they are considered more complex than storage-only tasks (Gathercole et al., 2004b; see also Vugs et al., 2013). However, it should be mentioned that given the dual nature of complex (span) WM tasks, some researchers have called for caution when interpreting the resources they target (Archibald & Harder Griebeling, 2016; Cowan et al., 2005).

A relatively recent addition to Baddeley’s original tripartite model is the episodic buffer (Baddeley, 2000), a passive temporary store capable of holding “unimodal or cross-modal inputs bound together into larger coherent chunks or episodes (e.g., sentences, stories)” (Montgomery et al., 2019, p. 3809f; see also Baddeley, 2003). It plays a crucial role in binding information (information integration) from the two subsidiary memory stores, i.e., the phonological loop and the visuospatial sketchpad, with information stored in long-term memory. The episodic buffer thus functions as an interface between working memory and long-term memory (Repovš & Baddeley, 2006). The episodic buffer is assumed to be controlled by the central executive, which can regularly retrieve information from it (Baddeley, 2000, p. 421). According to Alloway et al. (2004), the episodic buffer “may provide an important gateway for learning” (p. 87) and has been proposed to be “the source of capacity limits in language processing” (Boyle et al., 2013, p. 234) given its function as an interface between long-term memory and the other subsystems of WM. A graphic depiction of Baddeley’s multicomponent WM model is given in **Figure 1** below.

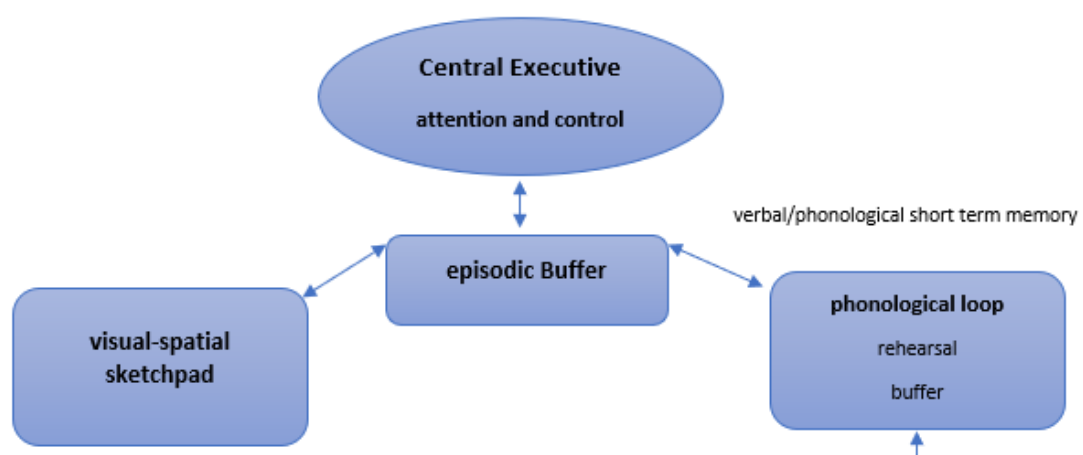


Figure 1: WM components (adopted from Baddeley et al., 2020)

WM abilities play an essential role in the development of language abilities (Archibald, 2018). It has been established that “the different subcomponents of the phonological loop function together to support different aspects of language learning and processing” (Pierce et al., 2017a, p. 1267f; see Gathercole, 2006b; Juffs & Harrington, 2011 for reviews). It has been argued that the phonological loop, specifically the phonological store⁵⁸, plays a crucial role in language acquisition by holding the incoming information stream in the memory long enough until it has been processed and transformed into LTM linguistic knowledge in the form of vocabulary or grammar (Baddeley et al., 2003; Pierce et al., 2017a). Accordingly, superior performance on nonword repetition tasks designed for assessment of VSTM has been associated with improved language learning abilities implying that “language acquisition is supported by the creation of adequate phonological representations within the phonological store” (Pierce et al., 2017a, p. 1268; see also Baddeley et al., 1998). Concerning oral language abilities, WM is highly implicated in sentence comprehension since it represents a complex problem-solving task (e.g., Ahmad Rusli & Montgomery, 2017). During comprehension, the listener is confronted with a rapidly decaying signal, based on which intermediate linguistic representations must be created and maintained in an active mental state while new ones are being established from the downstream material (Montgomery et al., 2018a). WM is further implicated in reading comprehension and writing abilities (Gathercole et al., 2004; Niedo et al., 2014) and is a significant predictor for longitudinal academic attainment (Gathercole et al., 2003). It has been argued that children with poorer WM capacities are particularly disadvantaged in structured learning activities involving intensive WM demands⁵⁹ (Gathercole & Alloway, 2008).

There is an inconsistency in research findings concerning bilingual WM advantage. On the one hand, many studies reported a bilingual advantage that even extends to bilingual children with low SES backgrounds (Blom et al., 2014; Brito et al., 2015, 2018; Delcenserie & Genesee, 2017; Kaushanskaya et al., 2014; Marini et al., 2019; Morales et al., 2013). In a meta-analysis featuring 27 independent studies (88 effect sizes), Grundy and Timmer (2017) found a significant small to medium overall effect size,

⁵⁸ The phonological store is argued to be implicated in language acquisition, but not the subvocal rehearsal, as there is ample evidence supporting the claim that phonological working memory is associated with language acquisition already in toddlerhood, while the subvocal rehearsal does not develop until age 7 or even later (Stokes & Klee, 2009; Pierce et al., 2017a).

implying greater working memory capacities in bilinguals relative to monolinguals. Furthermore, when different age groups were compared, the largest effect sizes emerged for bilingual children. On the other hand, several studies did not document a bilingual advantage for WM using similar tasks (e.g., Blom et al., 2017; Bonifacci et al., 2011; Engel de Abreu, 2011; Meir, 2017; Wimmer & Scherger, 2022). Some studies even report a bilingual disadvantage, particularly on verbal WM tasks. For example, while Luo et al. (2013) observed a superior performance of adult bilinguals on tasks tapping visuospatial working memory, they reported an inferior verbal WM performance of bilinguals relative to monolinguals on a word-span task. Messer et al. (2010) and Windsor et al. (2010) argue that the absence of bilingual advantage in verbal WM tasks could result from their lower linguistic skills. Some researchers have argued that VSTM and VWM tasks may be confounded by their verbal nature (Montgomery et al., 2010) and that the specific characteristics of the test stimuli in interaction with task demands might affect “different groups of learners differentially” (Pierce et al., 2017b). For example, Hamann (2017) provided an overview of studies demonstrating that performance on nonword repetition tasks incorporating test items with phonological or morphological resemblance to existing words in a particular language can be influenced by the extent to which individuals can draw on their LTM, i.e., their lexical knowledge and familiarity with the respective language. Indeed, sensitivity to different lexical (e.g., lexicality, frequency, concreteness, age of acquisition) and semantic properties (e.g., conceptual meaning, sense relations) of the to-be-remembered items has been shown to affect retrieval performance (Campoy et al., 2015; Conlin & Gathercole, 2006; Hargreaves et al., 2012; Jeffries et al., 2004; Lau et al., 2018; see Montgomery et al., 2019, p. 3812 for a discussion). This calls for caution when interpreting the performance of bilingual populations or socially-disadvantaged children on verbal WM tasks that draw on LTM linguistic knowledge like the word span and specific nonword repetition tasks (see also Hackman et al., 2014, for a discussion on the adverse effects of SES on WM). Moreover, when bilinguals are assessed in their weaker language using verbal WM tasks, they might be slower at lexical retrieval as a result of weaker lexical representations for the non-dominant language and interference from the dominant language (Ivanova & Costa, 2008; Kaushanskaya & Marian, 2007). This, in turn, might result in longer reaction times on verbal WM tasks.

Inhibition refers to the ability to resist interference or delay an automatic tendency towards a certain behavior, which is usually a previously learned response (Dempster, 1992). In studies on EFs, a distinction is typically made between interference inhibition

and response inhibition. *Interference inhibition* (also known as interference suppression) or control refers to the ability to resolve a perceptual conflict between competing cues by ignoring the more salient irrelevant stimulus and only attending to the relevant, less salient target cue (Martin-Rhee & Bialystok, 2008). *Response inhibition* (motor inhibition) involves the active suppression of a prepotent and habitual or familiar (automated) response and replacing it with an updated, less-automated response (de López & Baker, 2015) and can be further divided into action suppression and action cancellation (Schachar et al., 2007).

Inhibitory control (interference inhibition) is standardly assessed via tasks requiring the participant to disregard interference from irrelevant stimuli, such as the Simon task (Simon, 1969; see also Bialystok et al., 2004; Martin-Rhee & Bialystok, 2008), the Flanker task (Eriksen & Eriksen, 1974; see also Bunge et al., 2002; Carson & Meltzoff, 2008; Yang et al., 2011), and the Stroop task (Bialystok et al., 2008; Delis et al., 2001; Freeman et al., 2017), which consist of congruent (low-conflict) and incongruent (high-conflict) trials. In the incongruent trials, a cue interfering with the correct response is introduced, and the participants are supposed to suppress irrelevant visual information and respond only to the target feature. Examples of irrelevant information are “the position of the colored block in the Simon task, [...] arrows pointing in the opposite direction to the target arrow in the flanker task”, or names of colors presented in a different font color in the Stroop task (De Cat et al., 2018, p.122). Incongruent trials require a greater processing effort than congruent ones and tend to result in a slower response, i.e., longer reaction time. The reliable increase in response time from a congruent to an incongruent condition on the respective tasks is known as the Simon, Flanker, or Stroop effect (Dittrich et al., 2017; see review in Lu & Proctor, 1995), which is assumed to arise due to “conflict during response selection” (Martin-Rhee & Bialystok, 2008, p. 82). Costa et al. (2008) argue that tasks used to assess inhibitory control might implicate other EF components, particularly WM, to varying degrees. For instance, the Simon task counts as a complex response inhibition task since it is associated with moderate working memory demands. In this task, the participant is required to maintain “the stimulus-response rule in working memory (if it is red, press left; if it is blue, press right)” (Costa et al., 2008, p. 65) in addition to inhibition, whereas this working memory component is not necessary for the flanker task. A further recently developed task for evaluating information inhibition in children is the *Hidden Mouse* (Iluz-Cohen & Armon-Lotem, 2013, based on previous embedded figure tasks such as De Avila and

Duncan (1980). In this task, the child is presented with a series of pictures varying in degree of detail (i.e., in the amount of interference) in which a mouse is embedded and is requested to find the mouse as fast as possible. *Response inhibition*, on the other hand, is usually assessed using tasks involving conflict between “two response options to the same stimulus feature, creating a conflict between a habitual response and a less familiar arbitrary response” (Martin-Rhee & Bialystok, 2008, p.85). Examples of tasks assessing response inhibition are the VIMI (Verbal Inhibition, Motor Inhibition, Henry et al., 2012) task, the Go/No Go task (Donders, 1969), and the stop-signal paradigm (Logan, 1994). Such tasks require the participant to override the habitual response and replace it with a different one. Unlike interference suppression, this kind of inhibition mainly implicates the execution of motor responses and involves attentional control to a much lesser extent. Evidence from behavioral and neurolinguistic research has shown that bilingual advantages appear to depend on task demands and to selectively manifest in complex tasks that involve a high degree of interference and require control over attention to competing cues, i.e., in interference inhibition tasks but not in tasks solely assessing response inhibition (e.g., Carlson & Meltzoff, 2008; Luk et al., 2010; Martin-Rhee & Bialystok, 2008). A bilingual advantage for the former type of inhibition was observed for bilingual children at different ages, e.g., at age 5;0 (Martin-Rhee and Bialystok, 2008), age 5;6 (Morales et al., 2013), age 6;0 (Bialystok & Shapero, 2005; Bialystok & Viswanathan, 2009), age 7;0 (Poarch & van Hell, 2012), age 8;0 (Engel de Abreu et al., 2012) and in young adults as well (e.g., Qu et al., 2016). It should be mentioned, however, that some studies covering broader age ranges and using similar tasks did not find effects attributable to bilingualism (e.g., Duñabeitia et al., 2014; Gathercole et al., 2014; see Bialystok, 2017 for a discussion).

Inhibition has been one of the most extensively investigated EF components in research on bilingualism (see Hilchey & Klein, 2011, for a critical review). Motivated by the models put forward by Green (1998) and Miyake et al. (2000), early explanations of enhanced executive function in bilinguals were centered around the role inhibition plays in language control. It has been hypothesized in the latter models that the experience of suppressing the simultaneously activated competing information and allowing “processing to proceed through the target language had the long-term effect of boosting inhibitory processes more broadly” (Bialystok, 2017, p. 31). Despite its central role, several studies, in particular studies investigating bilingual infants and toddlers prior to the onset of productive language (e.g., Kovács & Mehler, 2009), have shown that an

explanation solely relying on inhibition “as the [...] single attentional control component in bilingual language control” (Poarch & Hell, 2012, p. 536) is not sufficient and research within the unity/diversity framework implies that inhibition is interrelated with and not truly separable from other EF-components (Miyake & Friedman, 2012). This resulted in a shift towards a more global executive functioning view that incorporates broader processes such as conflict monitoring (Botvinick et al., 2001; Costa et al., 2009; Hilchey & Klein, 2011), (selective) attention or coordination (Bialystok, 2011; Bialystok et al., 2012; Green & Abutalebi, 2013; Kroll & Bialystok, 2013; Kroll et al., 2012; see Valian, 2015 for an overview). Assuming that *attention* plays a crucial role in language selection, Engle and Kane (2004) introduced the notion of working memory capacity⁶⁰ as a continuous construct (also called “executive attention”, Engle 2002, 2018) as the combination of attention and WM, which is defined as “the cognitive system in which memory and attention interact to produce complex cognition” (Shipstead et al., 2015, p. 1863). *Attention* constitutes a multifaceted cognitive process that involves the ability to be alert, to selectively process a relevant stimulus amidst interference (selective attention) for a certain amount of time, and to sustain focus on the stream of information over time, i.e., sustain selective attention (sustained attention) (Montgomery et al., 2010; Posner & Petersen, 1990; Spaulding et al., 2008). Attention has been associated with language learning (see Ebert & Kohnert, 2011, for a review). For instance, attention may be required to draw focus to relevant linguistic stimuli prior to processing and to uphold this focus to avoid incomplete processing of that input. Furthermore, it has been argued that being able to rapidly engage and disengage attention is a prerequisite for processing rapidly presented input strings, which is a characteristic feature of linguistic input (Boerma et al., 2017). Although a bilingual advantage has been reported for different attention tasks, especially for selective attention (Blom et al., 2017; Engel de Abreu et al., 2012), studies examining bilingual effects yielded mixed findings for sustained attention (e.g., Bialystok et al., 2008; Boerma et al., 2017).

To summarize, despite the presence of a considerable amount of literature documenting bilingual cognitive advantages, a general EF advantage has been contested by some researchers (e.g., de Bruin et al., 2015a, 2015b; Dick et al., 2019; Duñabeitia et al., 2014; Hilchey & Klein, 2011; Hilchey et al., 2015; Paap & Greenberg, 2013; Papp et

⁶⁰ Note that the “capacity” in the “working memory capacity” construct does not denote storage space but rather refers to the “extent to which resources are available to control attention and maintain information relevant” for the task at hand (Bialystok, 2017, p. 34; Engle, 2002).

al., 2015, see Bialystok, 2017/2018; Leivada et al., 2021, Valian, 2015 for discussions). Moreover, some studies found bilingual effects only for specific tasks or under certain conditions, implying the existence of factors moderating the effect of bilingualism on EF (Boerma & Blom; Blom et al., 2017; Marton, 2016). In her review article, Bialystok (2017) links such conflicting findings to methodological issues, e.g., task features, statistical methods, sample characteristics, and classification of participants as bilinguals or monolinguals. Bialystok points out that most of the research reporting no differences between monolinguals and bilinguals was carried out with young adults using simple tasks (e.g., Gathercole et al., 2014; Kalia et al., 2014; Paap et al., 2015; Paap & Sawi, 2014) or covered relatively wide age ranges, e.g., Gathercole et al. (3-year-olds to older adults). Taking reaction times on the Flanker Task as an example, Bialystok (2017) explains that while children and older adults show slower reaction times, allowing for individual variation that could influence the outcome, young adults tend to perform at ceiling in such simple tasks. Consequently, Bialystok (2017, 2018) argues that group differences are unlikely to emerge given the insufficient variability at the inter-individual level and stresses the importance of operationalizing both “bilingualism” and “executive function” as a continuum rather than categorical variables, which should allow exploring the existence of possible linear relations “between the degree of bilingualism and the degree of cognitive function demands” (p. 283). For instance, some studies reported a cognitive advantage only for balanced bilinguals with high proficiency in both languages (e.g., Carlson & Meltzoff, 2008; Poarch & van Hell, 2012; Weber et al., 2016) or as an effect of increasing bilingual proficiency with more exposure to the respective language (e.g., Bialystok & Barac, 2012; Blom et al., 2014; Iluz-Cohen & Armon-Lotem, 2013). Some researchers even argue that language development supports EF development (Marcovitch & Zelazo, 2009) and that language abilities are determinants of children’s subsequent EF development (e.g., Bohlmann et al., 2015; Jones et al., 2020; Kuhn et al., 2016) suggesting that children with limited language ability are likely to have poor EFs. Hence, it is necessary to consider “the specific cognitive demands of executive function tasks” (Engel de Abreu et al., 2012, p. 1368) as well as factors contributing to the variability of the bilingual experience when exploring the impact of bilingualism on executive function (e.g., Abutalebi & Rietbergen, 2014; Kaushanskaya & Prior, 2015; Luk, 2015).

2.2 Specific Language Impairment (SLI)

Specific language impairment is a life-long neurodevelopmental disorder affecting the process of language acquisition in the absence of concomitant primary deficits, such as hearing impairment, impairments in neurological functioning, or socio-emotional and behavioral deficits indicative of autism spectrum disorders (Bishop, 1997, 2014; Leonard, 1998, 2014a). Children with SLI primarily exhibit deficits in language ability. Not only do they have delayed language development compared to unimpaired chronological-age peers, but they also show qualitative and quantitative differences in error patterns and rates (Leonard, 1998, 2014a, 2014b, 2014c; Rice, 2004; van der Lely, 1998). Importantly, the limited linguistic abilities cannot be attributed to “differences in the child’s cultural, linguistic or educational experiences” (Kohnert, 2013, p. 116) and are not part of a general cognitive delay⁶¹ or a biomedical condition⁶² (Bishop, 1997; Bishop et al., 2017; Clahsen, 1999; de Jong, 2008; Leonard, 2014a). With a prevalence rate of 7-10% among preschool children (see Norbury et al., 2016, for a UK-based study and Tomblin et al., 1997 for a US-based study), SLI constitutes one of the most common developmental communication disorders and is a public health concern given that affected children are at an increased risk of school failure and developing socio-emotional and behavioral problems (Law et al., 2013; Tomblin, 2008; Yew & O’Kearney, 2013).

SLI appears to run in families, as a substantial proportion of children with SLI are reported to exhibit a “family history of language difficulties, with a first-degree relative usually affected” (Conti-Ramsden & Durkin, 2012, p. 387). Studies of molecular genetics and familial aggregation (e.g., Choudhury & Benasich, 2003; O’Brien et al., 2003; Tomblin, 1989), as well as heritability estimates in twin studies (Bishop et al., 1995, 2006; Stromswold, 2001; Tomblin & Zang, 1999), indicate a strong genetic component to the disorder, which appears to affect more males than females (Conti-Ramsden & Botting, 1999; Flax et al., 2003). In a meta-analysis study, Stromswold (2001) pooled data from 7 studies and found an overall estimate of 86.6% concordance for monozygotic twins, compared to 50.2% for dizygotic twins. Moreover, siblings of individuals with language impairment were 30.3% more likely to be affected by the impairment than the general

⁶¹ At a group level, children with SLI tend to score below their typically-developing peers on measures for non-verbal IQ; nevertheless, they reliably perform above the level of intellectual disability (Leonard et al., 2021).

⁶² Conditions, where language impairment constitutes a part of a more complex pattern of impairment (Bishop et al., 2017).

population (see also Nudel et al., 2014; see Rice 2020b and Rice et al., 2020 for recent twin studies). It is, however, important to note that the genetic etiology of SLI is highly complex and multifactorial. Early studies within the genome project (Fisher et al., 1998; Hurst et al., 1990; Lai et al., 2001; SLI Consortium, 2002) suggested SLI to be associated with mutations on the *FOXP2* gene (based on a case study of a large 3-generation pedigree known as the ‘KE’ family). However, subsequent genetic studies have shown that *FOXP2* is only “one critical piece of a complex puzzle” (Fisher, 2005, p. 111) and that SLI is not caused by a variant or a mutation of a single gene (Bishop, 2002, 2006a, 2006b; Centanni et al., 2015) but rather involves contributions from multiple genes of small effect, “possibly in interaction with subtle environmental factors” (Bishop, 2006a; Leonard, 2014c, p.43; see Nudel et al., 2020 for a detailed overview). Although more recent research identified further candidate genes (e.g., *NFXL1*, *GNPTAB*, and very recently *BUD13*, see Andres et al., 2021), there remain no specific genetic markers that allow the identification of SLI through genetic screening (Misyak & Christiansen, 2011).

SLI is traditionally defined in terms of exclusionary criteria that help distinguish SLI from other clinical conditions. Leonard (2014a) provides a “summary of the areas of functioning that are evaluated prior to concluding that SLI is the most appropriate clinical label to apply” (p. 15).

Table 1: Criteria for SLI
(adopted from Leonard, 2014a, p. 15)

Factor	Criterion
Language ability	Scores indicative of language impairment on a test with satisfactory diagnostic accuracy and supportive evidence from independent judgment/clinical placement
Nonverbal IQ	Nonverbal IQ of at least 85 or a score that remains above 70 after standard error of measurement is considered
Hearing sensitivity	Pass screening at conventional levels
Otitis media with effusion	Unlikely to cause SLI by itself; nevertheless, for interpretation of language status, there should not be recent evidence of repeated episodes
Oral structure and function	No structural anomalies; pass screening using developmentally appropriate items and/or performance on oral-motor items clearly underestimates child’s observed expressive language
Social and physical interaction	No symptoms of impaired reciprocal social interaction or restriction of activities

Terminological remarks

Due to clinical considerations, some researchers (e.g., Reilly et al., 2014a) proposed abandoning the exclusionary criteria for SLI and dropping the defining label “specific” in order to facilitate access to speech and language intervention programs. This is motivated by concerns that “instead of being used for diagnostic differentiation, exclusionary criteria [are] sometimes interpreted as criteria for denying services to children” (Bishop et al., 2017, p. 1071). More recently, the international CATALISE Consortium (Bishop et al., 2017) has advocated using the more inclusive diagnostic label ‘Developmental Language Disorder’ (DLD) instead of SLI to refer to children who have language impairment that is not associated with a known biomedical condition, such as Down Syndrome, sensorineural hearing impairment/loss, and autism spectrum disorder. The newly coined diagnostic label “DLD” differs from SLI in that it also encompasses cases where language impairment co-occurs with other neurodevelopmental disorders like attention deficit hyperactivity disorder (ADHD) or impairments in the areas of motor coordination and speech (Volkers, 2018). Furthermore, DLD does not require a discrepancy between language skills and non-verbal cognitive ability, e.g., non-verbal IQ (Bishop, 2017; Bishop et al., 2017). Unlike SLI, an IQ score below 85⁶³ does not bar a diagnosis with DLD; hence, the distinction between specific and nonspecific language impairment (non-verbal IQ levels between 70 and 85) would no longer exist under the label ‘DLD’ (Rice, 2020a, 2020b). As such, children with SLI constitute a subset of DLD (see Rice, 2020a, and Volkens, 2018 for discussions). In a recent longitudinal twin study, Rice et al. (2020) observed a twinning effect for language but not for non-verbal cognition and found “heritability [to be] most consistent in the SLI group” (p. 793). Therefore, Rice (2020b) stresses the importance of maintaining the distinction between specific and nonspecific language impairment, especially in the age group 4-6 years.

Since the characteristics of the language-impaired children featured in this thesis are more compatible with the diagnostic label ‘SLI’, i.e., only children with a non-verbal IQ of 85 and above with no concomitant neurodevelopmental or motor coordination disorders were included in the studies, the label ‘SLI’ is used throughout this thesis, also for the sake of continuity with much of the existing literature on bilingual children with language impairment. It should, however, be noted that the participants’ characteristics also meet the criteria for DLD (cf. Bishop et al., 2017).

⁶³ DLD encompasses non-verbal IQ scores as low as 70 (Bishop et al., 2017).

2.2.1 Vulnerable domains in SLI

Children with SLI are characterized by great heterogeneity in terms of linguistic profiles and the severity of the disorder (Leonard, 2014a; Bishop et al., 2017). Depending on the severity of the deficits in comprehension abilities, some researchers, e.g., Conti-Ramsden and Botting (1999), identified two major clusters of SLI⁶⁴: children who show inferior expressive ability relative to language comprehension, i.e., expressive SLI, and children who show weaknesses in both production and comprehension, i.e., expressive-receptive or mixed SLI, which appears to prevail among children with SLI (Conti-Ramsden & Durkin, 2012). It should be noted that deficits in language comprehension, in particular, represent a serious risk factor for persistent language difficulties that are likely to continue into adolescence and adulthood (Bishop et al., 2012; Conti-Ramsden & Durkin, 2011; Nippold & Schwarz, 2002). On the other hand, cases where comprehension is impaired with intact expressive language abilities⁶⁵ are extremely rare among children with SLI.

As to the domains affected by the disorder, the areas of morphosyntax and phonology stand out. Studies on monolingual English-speaking children with SLI (Bishop et al., 2006; Falcato et al., 2008) provided evidence for the heritability of two classes of symptoms that serve as phenotypic markers for SLI. The first one involves aspects of grammatical computation⁶⁶, indicated by inconsistencies in using verbal inflections marking tense and agreement (Gopnik, 1990; Ullman & Gopnik, 1994), as well as difficulties in comprehension of syntactically complex constructions (Rice et al., 2009). The second weakness is argued to be related to deficits in “phonological short-term memory”, as reflected by poor performance in repeating nonwords that are three syllables or longer (Leonard, 2014a, p.19; see also Newbury et al., 2009). Although the latter two types of deficits often co-occur in children with SLI, evidence from the twin study by Bishop et al. (2006) has shown that, despite being significantly heritable, the two vulnerable areas were separable. It should also be noted that despite being most

⁶⁴ Note that the CATALISE consortium did not reach a consensus concerning the classification of DLD into subtypes. Instead of the DSM’s (Diagnostic and Statistical Manual of Mental Disorders) classical distinction between expressive and receptive language disorder, the panel recommended assessing domains of language that are considered to be most problematic for individuals with DLD (Bishop et al., 2017, p. 1073).

⁶⁵ Problems only in language comprehension occur frequently in children with autism spectrum disorder (Conti-Ramsden & Durkin, 2012).

⁶⁶ Linguistic operations such as recursion and hierarchical non-local dependencies between grammatical elements (van der Lely, 2005b, p. 53).

prominent, deficits in children with SLI are not restricted to grammatical computation and phonological short-term memory, as children with SLI frequently exhibit deficits in other areas, such as vocabulary, lexical semantics (Leonard, 2014a), as well as deficits in the interface between syntax-semantics and aspects of pragmatics⁶⁷ that primarily rely on levels of structural language knowledge (Andrés-Roqueta & Katsos, 2020; Bishop, 2000; Davies et al., 2016), e.g., referentiality (Fichman et al., 2022; Johnston et al., 1997; van der Lely, 1997a), universal quantification (Katsos et al., 2011), telicity and exhaustivity in *wh*-questions (Roeper, 2004; Schulz & Roeper, 2011). There is also increasing evidence pointing to the existence of subtypes of SLI, where particular domains of language are selectively impaired while others are spared (de Villiers, 2003; Friedmann & Novogrodsky, 2008; Novogrodsky, 2015), e.g., phonological SLI, grammatical or syntactic SLI (Friedmann & Novogrodsky, 2004; van der Lely, 2005a, 2005b), lexical SLI (Friedmann & Novogrodsky, 2011), and pragmatic SLI⁶⁸ (Bishop & Norbury, 2002; Botting & Conti-Ramsden, 2003). The subsequent sections provide a detailed description of vulnerable domains in SLI. Given the age range of the participants and the scope of the studies in this thesis, I refrain from elaborating on pragmatic deficits in SLI.

2.2.1.1 Morphosyntactic abilities of children with SLI

Although children with SLI usually demonstrate deficits in different language domains, such as phonology and lexicon, a large body of research has demonstrated that deficits in the comprehension and production of morphosyntax represent the hallmark of SLI (Hamann et al., 1998; Leonard, 2014a, 2014b; Marinis, 2011; van der Lely, 1998). Cross-linguistic research on SLI has shown that “in any given language, those features that pose a learning challenge for young typically developing children will prove to be an area of dramatic weakness in children with SLI” (Leonard, 2014b, p.2). However, the manifestation of such difficulties may vary depending on the typology of the language being acquired (Håkansson, 2017; Leonard, 2014b, 2015), so that clinical markers of SLI

⁶⁷ Note that aspects of pragmatic ability that are less reliant on formal/structural language knowledge, e.g., social pragmatic tasks requiring correct attributions of mental states (theory of mind -ToM, Baron-Cohen et al., 1985) appear to be relatively spared in children with SLI unlike their peers with autism spectrum disorders (ASD), who experience “exceptional difficulties with social pragmatic tasks” (Andrés-Roqueta & Katsos, 2020; see also Matthews et al., 2018 for a discussion on sources of individual variation in different areas pragmatic ability).

⁶⁸ Children with pragmatic impairment demonstrate deficits in social pragmatic comprehension without displaying the full spectrum of autism related deficits.

vary across languages, which could also overlap with errors of child bilinguals. Since two of the studies featured in this thesis are primarily concerned with sentence repetition (a measure of morphosyntax) as a diagnostic tool for SLI in bilingual contexts, the discussion on morphosyntactic abilities in SLI will be restricted to the domain of verbal morphology and complex syntactic constructions. Moreover, it has been shown in the preceding sections that aspects of nominal morphology are particularly subject to effects of the distributed nature of dual language exposure; hence, they are likely to represent areas of overlap between typical bilingual acquisition and aspects of impaired language acquisition in monolinguals.

Verbal inflectional morphology

In languages with sparse inflectional morphology, like English, children with SLI are reported to have exceptional difficulties with the acquisition of finite verb morphology⁶⁹ compared to nominal morphology, such as articles and plural inflections (Bedore & Leonard, 1998). This deficit in English SLI is also found under the label “grammatical tense marking” (Rice & Wexler, 1996; Rice et al., 2009) and is reported to have a strong hereditary component independent of non-word repetition (e.g., Bishop et al., 2006; Falcaro et al., 2008). Difficulties with tense and agreement morphology have also been reported for SLI children acquiring languages other than English, including Germanic languages like Danish (Lum & Bleses, 2012; Vang Christensen & Hansson, 2012), Dutch (Blom et al., 2013; de Jong, 1999; Rispens & De Bree, 2014; Verhoeven et al., 2011a; Wexler et al., 2004), German (Clahsen et al., 1997; Hamann et al., 1998; Rice et al., 1997; Rothweiler et al., 2012), Swedish (Hansson & Leonard, 2003; Leonard et al., 2004) and non-Germanic languages like Arabic (Abdalla, 2002, 2011), Hebrew (Dromi et al., 1993; Leonard & Dromi, 1994), Greek (Clahsen & Dalalakis, 1999; Stavrakaki et al., 2008), and Spanish (Bedore & Leonard, 2001; Grinstead et al., 2013, 2014) see Krok and Leonard (2015) for a meta-analysis.

⁶⁹ The exact nature of the deficit in finite verb morphology is subject to debate as there is no consensus on whether only tense (Tense Deficit Hypothesis: Rice et al., 1995; Rice & Wexler, 1996), only agreement (Agreement Deficit account: Clahsen, 1989; Clahsen et al., 1997) or even both are impaired, where both tense and/or agreement projections are optionally omitted (Agreement/Tense Omission Model: Wexler et al., 1998).

Within syntactic theory, main clauses, questions, as well as embedded clauses⁷⁰ can bear finiteness (cf. Chomsky, 1995; Haegemann, 1994; Rizzi, 1997; Quirk et al., 1985). According to Klein (2006), finiteness is what “links the descriptive content of [a] sentence (the ‘sentence basis’) to its topic component (in particular, to its topic time), and ... confines the illocutionary force to that topic component” (p. 245). For example, in declarative main clauses, finiteness indicates that an assertion, which is confined to a given time, i.e., the ‘topic time’, is made, where the tense component defines the relation between the topic time (reference time) and the utterance time (speech time) and “aspect the relation between reference time and event time” (Hamann et al., 2001, p. 194), see also von Stechow (1995). Given that Topichood is part of the CP features and projections, this implies that finiteness attracts movement, most probably to a projection within the Split CP⁷¹ (Rizzi, 1997). Klein (2006) further highlights the role of finiteness as an “organizing factor in the structure of an utterance” (p.249) as many structural phenomena (e.g., word order, movement, grammatical subject licensing, gapping, licensing of expletives) are tied to finiteness. This is “clearly reflected in the early stages of first and second language acquisition ... [in that] the acquisition of finiteness ... leads to a major restructuring of learner language” (Klein, 2006, p. 249; see also Klein & Perdue, 1997). Accordingly, morphology and syntax are closely related “because morphological elements carry word order and phrasal movement requirements, hence the term *morphosyntax*” (Rice et al., 2009, p. 1418; see Pollock, 1989, for details).

Overt Finiteness⁷² markers in English include the following set of morphemes marking the features tense (Tns) and sometimes agreement (Agr): third person singular non-past or habitual {-s} on lexical verbs, e.g., *she sings*, regular and irregular past tense on lexical verbs, e.g., *she danced/she sang*, copula and auxiliary *BE*, e.g., *she is smart/she is dancing*, and auxiliary *DO*, which co-occurs with the negation marker preceding lexical verbs in negated clauses, e.g., *she does not dance every day*. With respect to copula *BE*

⁷⁰ As part of the articulated Complementizer Phrase (Split CP), a special functional projection (FinP) has been proposed as the lowest projection within the CP-layer “... Force ... (Topic) ... (Focus) ... **Fin** IP” (Rizzi, 1997, p. 288). It specifies the finiteness of the embedded clause, i.e., whether the subordinate clause “bears finite or nonfinite tense” and links the IP-layer (the clause) to the superordinate CP-layer (matrix-clause or discourse, Rizzi, 1997).

⁷¹ Hamann et al. (2001) proposed anchoring tense into discourse via a specific projection for ‘topic time’ (reference time) within the CP, i.e., by placing a topic time operator into CP in order to bind the reference time argument that is located in TP.

⁷² Note that finiteness does not always feature verb marking and can be expressed differently depending on the typological properties of the language (cf. Håkansson, 2017).

and auxiliary *DO* forms in questions (e.g., copula *BE*: *Is the girl sad?/Why is the girl sad?*; auxiliary *DO*: *Does she usually visit her parents?/When does she visit her parents?*), they are said to “share the property of finiteness marking in a projection directly related to matrix clause structure” (Rice et al., 2009, p. 1421; see also Chomsky, 1995; Haegeman, 1994; Quirk et al., 1985). This is because during the derivation of questions from matrix clauses, inflectional material marking Tns and Agr features move to a projection higher than the inflectional phrase (IP), namely to the complementizer phrase (CP). The special status of auxiliary *Do* in English interrogatives (*do-support*) is related to the fact that, in contrast to other languages (e.g., V2 languages), thematic/lexical verb raising operates covertly in English (Pollock, 1989).

Deficits in using finite verb morphology repeatedly emerged as a reliable clinical marker for SLI in English-speaking children at least up to age 9;0 in simple declarative sentences (cf. Rice et al., 1995, 1998; Rice & Wexler, 1996, 2001) and have been shown to persist into adolescence in sentences involving CP projections (Rice et al., 2009). These deficits are captured in Rice and Wexler’s (1996) extended optional infinitive (EOI) account, which postulates that the optional infinitive period (OI, Wexler, 1994) observed in younger typically-developing children is prolonged in children with SLI. During the OI⁷³ stage, TD children (until approx. age 3) exhibit inconsistencies in using finiteness markings on the verb and alternate between correct tense marking and inappropriate production of non-finite verb forms in obligatory contexts, i.e., they tend to omit finiteness markings producing infinitive-like verb forms, e.g., *she laugh* instead of *she laughs*, while adhering to the relevant syntactic distributional restrictions associated with infinitives. Because optional infinitives appear to be restricted to main root clauses, they are also referred to as root infinitives (RI) (Hamann et al., 2003; Rizzi, 1994). Such an OI stage is not restricted to English and has been documented for many languages, such as Danish, Dutch, French, German, Norwegian, Swedish⁷⁴, Hebrew and Russian (Armon-Lotem, 1996; Bar-Shalom & Snyder, 1998; Blom, 2003; Brun et al., 1999; Haegemann, 1995; Hamann, 2002; Hamann & Plunkett, 1998; Kallestinova, 2007; Phillips, 1995, 2010; Platzack, 1990, 1992; Poeppel & Wexler, 1993; Rasetti, 1996, 2000; see Wexler,

⁷³ Note that slightly updated versions of the OI account exist in the literature, namely the Agreement Tense/Omission Model (Wexler et al., 1998), according to which root infinitives arise due to the optional projection of T/Agr features, and the Unique Checking Constraint Model (UCC; Rice et al., 1998; Wexler, 1998).

⁷⁴ Note that these languages differ from English in that infinitives bear an overt morphological marking, e.g., {-e} in Danish, {-en} in German, and {-er, -ir and -re} in French (Hamann, 2002, p. 168).

1994 for an overview), but are extremely rare in the early productions of children acquiring *pro*-drop languages like Catalan, Italian and Spanish (Guasti, 1994; Torrens, 1995; see Guasti, 2002 for a discussion; see Grinstead et al., 2014 for arguments pro the existence of an OI stage in early child Spanish).

Similar to typically developing children, English-speaking children with SLI are reported to omit finiteness markings along with a few commission errors (errors in form choice, e.g., *she are* instead of *she is*). However, they show much higher omission rates than younger language-matched TD peers (Cleave & Rice, 1997; Hoover et al., 2012; Leonard et al., 1997; Marshall & van der Lely, 2006), and their problems in this area continue until they reach the age of nine and beyond (Norbury et al., 2001; Rice et al., 1995, 1998, 2004). It should be noted that infinitival forms coincide with bare stems⁷⁵ in English. Leonard and Kueser (2019) argue that the predominant error type observed in English-speaking children with SLI, namely *omissions* of tense/agreement inflections, is likely to be tied to the abundance of bare stems in English, which coincide with infinitival forms. They further point out that this is not necessarily the case for other languages like German and Dutch and that other error types might predominate “in languages with few or no bare stems” (p. 351).

Compared to English, German has a more elaborate verbal inflection paradigm. For instance, in the present tense, English overtly marks subject-verb-agreement (SVA) only for third-person singular on main lexical verbs (e.g., *he talks*), whereas German possesses four overt inflectional forms marking person and number features: 1st PS SG {-e/Ø⁷⁶}, 2nd PS SG {-st}, 3rd PS SG and 2nd PS PL {-t} and 1st PS PL and 3rd PS PL {-en} (Scherger, 2019b). The latter agreement suffixes can be used with all tenses as well as with modal and auxiliary verbs, despite the existence of some suppletive SVA forms. Unlike English, the infinitival form in German consists of a stem and the suffix {-en}, which is usually replaced by an inflectional suffix in the case of finite verb forms. It should be noted that the German verbal agreement paradigm is characterized by syncretism, where the features ‘person’ and ‘number’ are expressed using the same inflectional affix, where agreement and tense are fused in the present tense (Eisenberg,

⁷⁵ The term “bare stem” is used to refer to a verb form consisting of the verbal stem and zero affixation (i.e., a null morpheme).

⁷⁶ Note that 1st PS SG marking is frequently omitted in colloquial German resulting in stem forms (Kauschke, 2012; Eisenberg 1999). In Southern German dialects, the infinitival form is marked using {-e}, which coincides with the inflectional suffix marking 1st PS SG in Standard German (Hamann, 2015).

2013). Hence, when overt finiteness marking is omitted (i.e., bare stem) or if the finite verb form is replaced by an infinitive, one cannot determine whether the tense or agreement feature is lacking (Hamann, 2015; Rice et al., 1997).

Concerning verb placement, German is argued to have an underlying OV order, where uninflected verb forms are base-generated within a head-final VP (Haider, 2010; Pollock, 1989), “or, given antisymmetry, it is assumed to have object movement to a case position outside of the VP” (Hamann, 2015, p. 219; cf. Kayne, 1994; Zwart, 1993), as a result of which nonfinite verbs occur clause-finally with their complements to their left as exemplified in (8) below (adapted from Hamann, 2015).

- (8) Laura möchte morgen **ein Auto kaufen**
 Laura want-3SG tomorrow a car buy-*inf*.
 ‘Laura wants to buy a car tomorrow’

An important characteristic of German is that it is a verb-second language⁷⁷, in which finite verbs appear in second position (V2) in main (root) clauses (both subject-initial and non-subject-initial) and in final position in embedded clauses⁷⁸ (Hamann, 2015). Verb movement to V2 is triggered by finiteness in that the verb moves from V to I and subsequently from I to C leading to the so-called V2 phenomenon⁷⁹ in main clauses (Hamann, 2015, p. 219). Non-finite verb forms (e.g., *en*-infinitives and participial forms) and verb stems are not licensed in V2 (i.e., not licensed in C⁰ and are restricted to V⁰); see

⁷⁷ In V2 languages, constituents other than the subject can occupy sentence initial position, i.e., linearly precede the finite verb in main clauses, e.g., structures with topicalized and focalized elements.

⁷⁸ Following traditional analysis of German syntax (e.g., Grewendorf, 1988), in embedded (subordinate) clauses, the lexical verb is raised from V⁰ to I⁰ in a head-final IP to check for finiteness features. However, further movement to the CP layer is blocked since C⁰ is already occupied by a complementizer or a similar element. As a result, the finite verb stays in-situ in the IP head position and appears clause-finally in embedded clauses, e.g., *dass der Sohn den Vater besucht* ‘that the.NOM son visits the.ACC father’ [_{CP} dass [_{IP} der Sohn [_{VP} den Vater *t_i*] besucht_i]].

⁷⁹ It has been proposed that German clauses, in which the subject occupies the first position, do not necessarily involve a CP projection and are rather IP or AgrP/TP projections (residual V2, Hamann, 2015; see von Stechow & Sternefeld, 1988 for a discussion). Similarly, adverbs have been argued to occupy a different position than fronted objects (focused or topicalized), see Bühring (1995) for semantic/pragmatic arguments and Rizzi (2004) for syntactic ones. Hence, it can be presumed that only those structures featuring fronted objects can be truly indicative of the acquisition of a ‘generalized V2’ (Hamann, 2015). However, several studies do not make a distinction between subject and non-subject-initial clauses and use ‘generalized V2’ to refer to an acquisitional phase, in which finite verbs are placed in second position including subject-initial and AdvP initial V2 clauses (e.g., Clahsen, 1999), i.e., for clauses where “one single constituent linearly precedes the finite verb” (Schulz & Schwarze, 2017, p. 55).

Holmberg (2015) for an overview of V2; see also Vikner (1995). The combination of the OV and V2 properties gives rise to the so-called sentence bracket ‘Satzklammer’ in German main clauses. The sentence bracket is a term from traditional grammar according to which the middle field of a German sentence is bracketed by the finite auxiliary or modal and the non-finite lexical verb, i.e., the finite auxiliary or modal verb occupies V2, whereas the non-finite verb occupies the sentence-final position. In Generative Grammar, this property is explained by movement of the finite verb (see Hamann et al., 2017 for details). Particle verbs constitute a special case of the German sentence bracket as the finite verb moves to C⁰ (hence appears in V2) leaving the particle in the base position in the VP (Hamann, 2015; cf. Koster, 1975), as exemplified in (9) below.

- (9) Tom macht das Fenster im Wohnzimmer **zu**
 Tom makes the window in the living room closed
 ‘Tom closes the window in the living room’

The acquisition of SVA and the correct placement of finite verbs in V2 go hand in hand in German and belong to the early acquired phenomena (Grimm & Schulz, 2016; Penner et al., 1994; Tracy, 1991). Typically developing children usually achieve target-like mastery of the verbal agreement paradigm by the end of their third year of life (Clahsen et al., 1996a; Clahsen & Penke, 1992; Hamann, 2015; Siegmüller, 2014). Before full mastery of the paradigm, TD children go through a phase during which they optionally use infinitival verb forms side by side with inflected verb forms⁸⁰ (Clahsen 1986a, 1986b; Poeppel & Wexler, 1993; Weissenborn, 1990; Wexler, 1994). The first agreement suffix to emerge is 3PS SG {-t}, whereas the last one to be acquired is 2PS SG

⁸⁰ Several theoretical accounts have been put forward to explain the phases monolingual TD children go through until they master SVA and V2 placement, e.g., Poeppel and Wexler’s (1993) Optional Infinitive account, Rizzi’s (1994) Truncation Hypothesis and Clahsen et al.’s (1996) Lexical Learning Hypothesis. However, an elaborate discussion of these accounts is beyond the scope of this thesis, see Hamann (2002) for an in-depth discussion.

{-st}⁸¹ (Clahsen, 1988, 1991a; Clahsen et al., 1996a; Mills, 1985; Siegmüller, 2014). Frequently attested error patterns during this period include substitutions of the agreement markers by the infinitival inflectional marking {-en} and omissions of inflectional suffixes, i.e., production of bare (-Ø) stems⁸² (e.g., *da hang-Ø das* ‘there hang this’, MoTD, age 1; 11, Clahsen et al., 1993, p. 411). By contrast, incorrect application of suffixes not agreeing with the subject in the ‘person’ or ‘number’ features is rather rare in TD children.

As to verb placement, TD children are reported to show mastery of the German OV order already during the two-word stage (Penner et al., 1994). Crucially, they appear to obey “the ban on non-finite verbs in V2” (Schulz & Schwarze, 2017, p. 52) during the optional infinitives stage. Though finite verbs are sometimes placed in sentence-final position (Clahsen et al., 1996b; Penner, 1990) as a reflex of the basic OV order in German and probably as a truncation option, TD children typically fluctuate between target-like placement of finite verbs in V2 and non-target-like production of infinitives in final positions in main clauses (root infinitives⁸³), e.g., *ich der Fos hab-n* ‘I the frog have’ (MoTD, age: 2;1, Wexler, 1994, p. 315). In contrast, non-finite forms, particularly -en infinitives, very rarely occur in V2 positions (Clahsen & Penke, 1992; Poepl & Wexler, 1993; Tracy, 1991; Weissenborn, 1990). This pattern implies that monolingual TD

⁸¹ Grijzenhout and Penke (2005) point out that phonological and prosodic developmental factors are likely to be behind the relatively later emergence of the 2PS SG {-st} inflectional suffix in German since this type of inflection results in a complex coda cluster. This is also in line with Demuth’s (2014, 2019) *Prosodic Licensing Hypothesis*, according to which a child’s ability to produce (complex) grammatical morphemes increases as her “prosodic (and more general phonological) competence develops” (Demuth, 2014, p.91). This is also assumed to interact with the prosodic properties of the language being acquired (Demuth & Tremblay, 2008; see also Gwinner et al., 2012).

⁸² Note that -Ø stems have been treated differently across studies. For example, Haznedar and Schwartz (1997), Prévost (2003), Schulz and Schwarze (2017), Schwarze et al. (2015), and Wojtecka et al. (2013) coded -Ø stems in V2 as finite, Rothweiler et al. (2012, 2017) coded them as “potentially non-finite”, while other studies like Clahsen et al. (1997) excluded them from the analyses altogether due to their ambiguous status.

⁸³ Note that root infinitives are allowed in certain pragmatic contexts in German, as in in Orders (*Aufstehen, bitte!* ‘Stand up, please’), instructions (e.g., *Die Pizza vorsichtig schneiden*, ‘cut the Pizza carefully’), granting permissions (e.g., *Ruhig fragen, wenn etwas unklar ist* ‘simply ask in case of unclarities’), and wishes (*Hurra! Endlich nach Mallorca fahren!* ‘finally travel to Mallorca’), see Gärtner, 2014).

children are sensitive to the relation between finiteness and verb-second placement⁸⁴ from early on (Clahsen & Penke, 1992; Poeppel & Wexler, 1993; Schulz & Schwarze, 2017). In other words, “even if the use of an infinitive is possible in the child grammar, this grammar already has a position for inflectional material and contains a verb raising rule in those cases where the target language has one” (Hamann, 2002, p. 170).

In many early languages, including German, children have been reported to optionally drop subjects in verbal utterances as long as they still use optional infinitives (Haegeman, 1995; Hamann & Plunkett, 1998; Hyams, 1986; Phillips, 1995; Rasetti, 2000), which represents “a rather marked stage in German” (Hamann, 2015 p. 220; see Hamann, 1996, 2002; Weissenborn, 1991)⁸⁵. Although null subjects tend to occur more frequently in root infinitives, they also occur in finite verbal utterances. Concerning subject omissions by verb type, results are mixed as some children show higher rates of null subjects with main lexical verbs than with copulas and auxiliaries (e.g., Hamann & Plunkett, 1998), whereas others display the reverse pattern (Hamann et al., 1996; Rasetti, 2000). Hamann and Plunkett (1998) analyzed longitudinal data obtained from two monolingual children acquiring Danish⁸⁶ over the period of five years (Jens: age span 1;00;2-6;1;2 and Anne: age span 1;01;1-5;10;22) for “subject omission, finite and infinitival verb usage and discourse anchorage of sentence subjects (overt and missing)” (p. 35). Their results show that, for a period of 12 months, Jens omitted subjects in more

⁸⁴ In spite of the consensus on the close temporal relation between the acquisition of SVA and verb-second placement in main clauses in German typical language development, the directionality of this relationship is still debated (Kauschke, 2012; Thelen, 2014; Weissenborn, 2000). According to Clahsen (1986a, 1991b) and Clahsen et al. (1997), the full acquisition of the SVA morphological paradigm and the C-projection is a prerequisite for triggering verb movement to V2. Weissenborn (1990) and Verrips and Weissenborn (1992); however, argue that finiteness and not acquisition of the full morphological paradigm of SVA per se is what triggers verb movement and the acquisition of V2-placement. They further assume that, even before SVA is fully acquired, children would project a fully-specified CP from early on, in which finite verbs occupy ‘C’. In this respect, Grimm (1993), Penner (1995a), and Penner and Kölliker Funk (1998) argue that children need to first master V2 placement in order to acquire the verbal agreement paradigm successfully.

⁸⁵ Note that “[a]dult German and Dutch are topic-drop languages”, but only clause-initially and not in wh-questions or embedded clauses (Guasti, 2002, p. 164). However, the null-subject pattern observed in early languages does not represent an option in adult grammar (Guasti, 2002). Moreover, null subjects are attested in early verbal utterances of children acquiring non-topic-drop languages like Danish and are thus unlikely to be input driven (Hamann & Plunkett, 1988).

⁸⁶ Although Danish is V2 language like German and Dutch with unambiguous morphological marking on infinitives, it resembles English in that it is an SVO language and lacks V-to-I raising. Moreover, it does not have overt morphological agreement marking and uses the inflectional suffix {-er} to mark finiteness for all types of grammatical person (Hamann, 2002, p. 151).

than 30% of the obligatory contexts on average, with a peak of 65% at age 2;3, whereas Anne demonstrated an omission rate of about 25%, with a peak of 40% at age 2;0. At ages 2;7 and 2;4, respectively, a drastic decrease in subject omission rates (< 5% omission rate) was observed for both children, finally approaching adult-like performance. Importantly, Hamann and Plunkett found that there was no phase in early Danish in which children exclusively used infinitives and that null subjects occurred in both finite and infinitival verbal utterances with almost equal distributions when omission rates were at their highest. Moreover, a comparison of the developmental profiles of root infinitives and overall null-subject rates in finite clauses revealed a parallel development between the two with a strong significant correlation ($r = .90$, $p < .001$): “infinitives and missing subjects appear together, reach their maxima at the same time, and go into a steep decline at the same time” (Hamann & Plunkett, 1998, p. 59; see Rasetti, 2000 for similar findings on French). Furthermore, patterns of placement of negation in relation to finite verbs corroborate the assumption that, despite being in the OI stage, children are “aware of the syntactic consequences of the finiteness marking” (Hamann & Plunkett, 1998, p. 66). Crucially, the fact that children use finite verbal utterances alongside root infinitives from early on and that null subjects also occur with finite verbs challenges the structure-building account (Radford, 1990). The latter account assumes the existence of an early phase in grammatical development, in which children exclusively produce infinitival verbal clauses lacking functional categories that can host inflectional material leading to subject omission.

By the time TD German-speaking children show mastery of the verbal inflectional paradigm and have a fully developed clause structure, children with SLI would still exhibit difficulties with the verbal inflectional system and the correct placement of finite and nonfinite verb forms, where some children with SLI continue to produce non-target SVA in main clauses and in more complex constructions up to school age and beyond (Clahsen, 1991a; Hamann et al., 1998, 2001; Lemmer, 2018; Lindner, 2002; Rothweiler et al., 2012, 2017; Schulz & Schwarze, 2017; see Hamann, 2015 & Leonard, 2015 for overviews). Apart from prolonged use of inappropriate nonfinite verb forms in utterance final position in declaratives (which is also attested in early languages of younger typically developing peers but much less frequently), German children with SLI (monolingual and bilingual) have been reported to produce non-finite verb forms in V2 position, a pattern which is very rarely attested in monolingual typical development and has thus been argued to be characteristic of German SLI (Chilla, 2008; Clahsen 1991a;

Clahsen et al., 1997; Rice et al., 1997; Schulz & Schwarze, 2017; Schwarze et al., 2015). Moreover, some studies reported a tendency for some children with SLI to place finite verbs in clause-final position in main clauses and even in *wh*-questions (Bartke, 1994; Grimm, 1986, 1993; Grimm & Weinert, 1990; Hamann et al., 1998; Kaltenbacher & Lindner, 1990). The latter error type has also been reported for younger TD children but to a much lesser extent (Clahsen et al., 1996b; Penner, 1990).

Several studies examined verbal inflectional morphology in monolingual German-speaking children with SLI and reported subject-verb-agreement rather than tense marking to be the most dominant error type in German SLI among finiteness markings (Bartke, 1998; Clahsen, 1991a; Clahsen et al., 1997; Rothweiler & Clahsen, 1993; Rothweiler et al., 2012, 2017⁸⁷). Clahsen (1991) analyzed spontaneous speech data of 10 monolingual German-speaking children with SLI aged between 3;2 and 9;6 (mean age = 5;6). His analyses showed overall low rates of target-consistent overt SVA marking and frequent omissions of auxiliary verbs. During the period of observation, the majority of the SLI children (9/10) did not achieve mastery of the full verbal agreement paradigm and tended to use a limited set of suffixes, where the 2nd PS SG suffix {-st} appeared to be particularly vulnerable. Moreover, the only child who managed to acquire the SVA inflectional paradigm achieved this considerably later than what has been reported in the literature for monolingual TD children (at around age 4 versus 3;0-3;6), see Clahsen et al. (1996a); Clahsen and Penke (1992). Frequent error patterns included using zero affixation (i.e., bare stems) and infinitival verb forms (i.e., *-n* forms) regardless of the sentential subject's grammatical person or number. Concerning verb placement in relation to finiteness, Clahsen's results showed that nonfinite verb forms (bare stems and infinitives) were primarily placed in final position in main clauses, whereas finite verb forms (predominantly *-t* forms) appeared in V2. His data also showed some individual variation as two of the SLI children placed both finite and nonfinite verbs clause-finally, and some children produced questions with the finite verb in situ, e.g., *Uschi hier wohnt?* 'Uschi here lives', [Does Uschi live here?] (MoSLI, 4;5; Clahsen, 1991a, p. 192). A similar pattern was observed in Rothweiler and Clahsen's (1993) longitudinal study, which examined SVA in 19 monolingual German-speaking children with SLI (age range: 3;11-6;11; mean age: 5;4) over the period of 12 months. During that period, only 2 out of

⁸⁷ Note that these studies were conducted on spontaneous speech samples from overlapping samples of monolingual children with SLI.

19 children with SLI managed to acquire SVA, yet with significant delays. Eleven out of 19 MoSLI children demonstrated lower accuracy scores on SVA than their younger MLU-matched typically developing peers. The latter SLI children used the 2nd PS SG {-st} suffix in less than 50% of the obligatory contexts and overused *-n* infinitives as a default form for replacing finite verb forms (mean rate of inappropriate infinitive use = 60%), see Bartke (1998) and Kaltenbacher and Lindner (1990) for similar findings on SVA in German-speaking MoSLI children. Participial markings, on the other hand, appeared to be relatively spared (Clahsen & Rothweiler, 1993; Rothweiler & Clahsen, 1993) and were correctly marked in 91.2% of the cases, presumably because participles do not display grammatical agreement (Rothweiler & Clahsen, 1993), see Kauschke et al. (2017) for contrasting findings on participial marking using elicited production. Based on their findings, Clahsen and collaborators argued that difficulties with agreement [but not with tense] are at the core of the grammatical deficit in SLI (Clahsen, 1999), which they coined under the *Agreement Deficit Hypothesis (ADH)*. The ADH was initially formulated within the framework of Generalized Phrase Structure Grammar as *the Missing Agreement Hypothesis* (Clahsen, 1988). According to it, children with SLI have a selective deficit in establishing grammatical agreement relations “between different elements of a phrase or a clause”, as in subject-verb-agreement, structural case, and grammatical gender marking on determiners and adjectives (Hamann et al., 1998, p. 201). Specifically, Clahsen proposed that the control-agreement principle (CAP; Gazdar et al., 1985), which determines the choice of grammatical features, “is selectively impaired in SLI, leaving configurational principles intact” (Clahsen et al., 1997, p. 153). Confining this to SVA, children with SLI are predicted to fail in spelling out the features ‘person’ and ‘number’, which are “controlled by the subject in subject-verb configurations” (Hamann et al., 1998, p. 201; Clahsen, 1991b). Hence, an asymmetry is to be expected between tense and agreement marking, where tense but not agreement features would be spelled out resulting in agreement errors, such as incorrect marking of grammatical person, cooccurrence of infinitives with overt subjects, and overuse of infinitival forms in clause-final positions. Concerning verb placement in main clauses, finite verbs should appear in V2 and nonfinite verb forms in final position (see also Hansen, 1994). Subsequently, the *Agreement Deficit Account* was rephrased within the minimalist framework combining it with Chomsky’s (1995) theory of formal features (Clahsen et al., 1997). Within this model, “the finite verb first mov[es] to Agr/T to check tense and agreement features and then [to C in order to] check[...] a feature related to finiteness

located in C” giving rise to the V2 phenomenon (Hamann, 2015, p. 225). What is crucial in this model is the notion of feature checking and that it distinguishes between intrinsic and optional features, both of which can be either interpretable⁸⁸ or non-interpretable (Clahsen et al., 1997, p. 153; see Chomsky, 1995, Chapter 4 for an overview). While interpretable features are not required to enter checking relations, non-interpretable features must be checked and eliminated through the course of derivation “so that only interpretable features reach the interpretive component and the C-I interface” (Hamann, 2015, p. 225). For instance, ϕ -features of verbs (and adjectives), such as ‘person’ and ‘gender’, constitute a class of optional non-interpretable features that must be checked off before LF. Since Clahsen and collaborators advocate for a selective deficit affecting the agreement but not the tense feature, they adopt the narrower view⁸⁹ of the Agreement deficit Hypothesis, according to which only the optional non-interpretable phi-features of verbs, namely the verbal agreement features ‘person’ and ‘number’, are affected in SLI child grammar. When the latter features are left unchecked, non-finite verb forms surface. Note that under the narrower interpretation of the ADH, case features (structural case), including nominative assignment by T and phi-features on subjects, would be spared even in sentences in which verbal agreement features are left unspecified (Clahsen et al., 1997, p. 154). A further important prediction is that verbs should be able to raise regardless of their feature specification since only strong features in functional heads trigger verb movement, and the strength of the categorical features does not depend on the verb’s phi-features. Hamann (2015) argues that the latter prediction is rather problematic, especially for the present tense, where finiteness is implicitly marked “through the presence of person and number morphology” (p. 226).

To support the narrower version of the ADH, which maintains that phi-features on verbs (person and number) but not tense are impaired in SLI, Clahsen et al. (1997) compared German and English SLI data. For German, Clahsen et al. analyzed 13 sixty-minute recorded sessions of spontaneous speech data from 6 children with SLI (age range: 5;8-7;11, MLU range: 2;6-4;2). For English, they analyzed elicited production longitudinal data from 9 English-speaking children with SLI (age range: 10;0-13;01). The

⁸⁸ Interpretable features are relevant for semantic interpretation, whereas non-interpretable features are not (Chomsky, 1995).

⁸⁹ The other broader reinterpretation of the ADH stipulates, analog to the CAP, that all non-interpretable features (case features, ϕ -features of adjectives and tense features of verbs) are affected in SLI (Clahsen, 1997, p. 153, see also Tsimplí & Mastropavlou, 2008).

English data was originally investigated by van der Lely and Stollwerck (1997) and consisted of recordings of two elicited production tasks targeting present 3rd PS SG and past tense forms, which were repeated within a 12-month interval. Contrary to Rice et al. (1995, 1996), who postulate that the deficit in SLI concerns tense and hence predicts English-speaking children to encounter difficulties with past tense morphology, the ADH predicts that SLI children with L1 English would specifically have problems with 3rd PS SG marking {-s} as this suffix is controlled by the phi-features of the verb (Clahsen et al., 1997, p. 155). Accuracy scores revealed a significant difference between the use of tense and agreement markings. As to main lexical verbs, the past tense {-ed} was correctly used in 75% of the obligatory contexts (N = 278) at T1 and in 78% of the obligatory contexts at T2 (N = 129). On the other hand, third person singular {-s} was correctly used in 40% of the obligatory contexts (N = 97 contexts) at T1 and 48% at T2 (N = 63). A similar pattern is reported for auxiliaries, where tense was correctly marked on 89% of the auxiliaries (137/154), whereas agreement was only correctly marked in 35% of the auxiliaries (7/20). Given the significant differences between tense and agreement markings, Clahsen et al. (1997) conclude that the core deficit in English SLI concerns agreement rather than tense. It should be noted, however, that the number of obligatory contexts for tense markings on auxiliaries was almost six times higher than that for agreement marking, which could have led to statistical confounds. Hamann (2015) highlighted methodological issues that are likely to be responsible for the reverse pattern (tense spared but impaired SVA) reported in Clahsen et al. (1997). First, the inflectional suffix {-s} not only marks person but also present tense and is thus always considered in counts of finite verb forms, especially after the EOI was rephrased via the Unique Checking Constraint⁹⁰, e.g., Rice and Wexler (1996) and Paradis and Crago (2001), see also van der Lely (1998). Moreover, as van der Lely (1998) pointed out, the data on tense were obtained from elicited narratives, whereas the data on agreement stemmed from a structured SVA elicitation task, resulting in a disproportionately high number of obligatory contexts for tense in the data analyzed by Clahsen et al. (1997). According to van der Lely, the same children had no more than 20% accuracy scores on

⁹⁰ The Unique Checking Constraint (UCC) is assumed to constrain children's syntactic development in that children's grammars allow the subject to check the uninterpretable D-feature of either tense or agreement, but not both. From an OT perspective (Prince & Smolensky, 2004), the competition between the UCC and the Minimize Violations constraint, which requires both of the D-features T and Agr to be checked as in the case of adult grammar, gives rise to the optional behavior in finiteness marking in early languages, depending on what constraint is being followed, see Chapter 4 of Guasti (2002) for a discussion.

a similar structured elicitation task targeting past tense, implying that tense and not agreement is the locus of difficulty in English SLI.

With respect to nominative case marking (subject case), the ADH predicts that SLI children should not encounter particular problems since nominative case is assigned by the functional category T, which is said to have an intrinsic case feature (Chomsky 1995). Clahsen et al. (1997) assume that phi-features of the verb are “independent of the case feature of T and that T is not affected in SLI”. Accordingly, nominative assignment should be possible “even in cases in which subject-verb agreement breaks down” (Clahsen et al., 1997, p. 159). Their results show that, unlike TD data, in which finiteness and nominative case were correlated (non-nominative subjects primarily appeared together with uninflected verb forms), SLI children produced a total of 217 nominative subjects and did not produce any non-nominative subjects “even in sentences with uninflected verb forms” (Clahsen et al., 1997, p. 158). In this respect, Hamann (2015) argues that the SLI children analyzed by Clahsen et al. (1997) are comparatively older, and the absence of non-nominative subjects in the data is likely to be an effect of extensive speech-language therapy; see Wexler et al. (1998) for contrasting findings on non-nominative markings in younger children with SLI.

Analyses of the German data showed that accuracy scores for SVA on lexical and auxiliary verbs (64% and 89% correct, respectively) were significantly lower than for past tense (preterite suffix), where children performed at ceiling on both lexical and auxiliary verbs. Moreover, an inspection of error patterns revealed that only a small proportion of the verb forms with incorrect SVA (31/313) were *-n* infinitives, whereas the vast majority of the errors were cases of incorrect agreement, where finite verb forms with 1st, 2nd, or 3rd PS SG or a finite form of *SEIN* were used incorrectly. Even after excluding ambiguous verb forms, i.e., verb forms with zero affixation, a total of 97 cases of agreement errors could still be found (ca. 31% of all forms with incorrect SVA). With respect to verb placement in relation to finiteness, at a group level, SLI children appeared to resemble their TD peers in that finite verbs were predominantly placed in V2 (87% of the finite verbs), while 89.5% of the non-finite verb forms occurred clause-finally. The sentence bracket was also respected in 96% of the cases, but in 4% of the cases, the finite verb linearly followed the non-finite verb at clause-final position. However, there was one SLI child (Dieter) who placed non-finite verb forms in V2 (11 out of 48 non-finite verb forms), a pattern that is almost non-existent in the speech of typically developing children (Clahsen et al., 1996a). Moreover, two further children showed a tendency to place finite

verbs in clause-final position, where 25% and 46% of their finite verbs appeared in clause-final position. The latter two children also extended this pattern to sentences with the sentence bracket. Placing finite verbs clause-finally is also attested in the speech of unimpaired children during the root infinitive stage, but to a much lesser extent (7-20%; see Clahsen et al., 1996a, Penner, 1990). To account for the occurrence of some of the finite verbs in clause-final position, Clahsen et al. (1996a) argue that these children sometimes do not project a CP, which carries strong categorical features that would attract the finite verb. As a result, the finite verb is not raised to C and remains in clause-final position. To explain the occurrence of Dieter's infinitives in V2, Clahsen et al. assume that the child projects a CP for each sentence that includes a verb, which is assumed to have a strong C feature that needs to be checked via verb raising. They further argue that "[n]on-finite verbs can be raised to Comp, as these verbs are unspecified for ϕ -features and do not have any features that would produce a feature mismatch in Comp" (Clahsen et al., 1997, p. 165). Despite the individual variation observed in 3 out of 6 of the examined SLI children, Clahsen et al. (1997) still conclude that all of the investigated SLI children have acquired V2 and argue for a selective deficit in verbal agreement in SLI that can be extended to English SLI data. Additionally, they argue that "verb placement and verb inflection are linked in SLI grammars in similar ways as in the grammar of unimpaired children" (p. 165).

Further support for a selective verbal agreement deficit in German SLI came from subsequent studies featuring Clahsen's (1991a, 1991b) monolingual sample with SLI, e.g., Clahsen et al. (2014), Rothweiler et al. (2012, 2017), and Ruberg et al. (2019). Rothweiler et al. (2012, 2017) and Ruberg et al. (2019) compared monolingual German-speaking children with SLI to bilingual German-Turkish children with and without SLI (overall age range: 3;6-7;11). The studies reported persistent problems with SVA for both SLI groups, even in language-impaired children who produced grammatically well-formed complex constructions like *wh*-questions and subordinate clauses. Children with SLI significantly produced more uninflected verb forms than TD children, which mainly appeared in clause-final positions, in addition to genuine agreement errors (commission errors)⁹¹, e.g., *dann macht ich diese*, 'then I make 3rd PS SG these', Rothweiler et al., 2012, p. 47). The studies also replicated Clahsen et al.'s (1997) findings on verb

⁹¹ Schulz and Schwarze (2017) recalculated the frequencies of different error types in the latter studies and found that bare stems (17%) and *-en* infinitives (6%) occurred more frequently than substitutions, with an error rate of only 3%.

placement with respect to finiteness and found V2 movement to be largely limited to finite verb forms, whereas non-finite verb forms mainly occurred clause-finally and rarely in V2. Note that bare stems were treated as potentially non-finite in the latter studies. Consequently, the authors argue against a broader language deficit and claim that, at least for the children they investigated, “syntactic computation and representation involving the CP-domain is not particularly affected” (Ruberg et al., 2019, p. 9). Instead, they propose SVA to be a reliable clinical marker for identifying SLI in monolingual and bilingual children.

Rice et al. (1995, 1997) challenged the Agreement Deficit Hypothesis and approached grammatical limitations in German-speaking children with SLI from an EOI perspective (Wexler et al., 1997), arriving at different findings from those of Clahsen and collaborators. In their longitudinal study, Rice et al. (1997) compared spontaneous speech data of a group of 8 young monolingual German-speaking children with SLI (ages 4;0 to 4;8, mean MLU: 2.76) to a control group of 8 younger MLU-matched TD children (ages 2;1 to 2;7, mean MLU: 2.76) at two points of time (T1 & T2) with an interval of 12 months. In their analyses, Rice et al. assume that an inflected verb in German is marked for finiteness, i.e., for both features Tns and Agr, since the features Tns and Agr are “fused in the present tense” and it cannot be assured which feature of the two is missing if an infinitival form is used or when auxiliary or copula verbs are omitted (Hamann, 2015, p. 229). Similar to their English peers with SLI, Rice et al. anticipated German children with SLI to be more likely to omit copular *SEIN* (as a finiteness marker) and to use infinitival forms of lexical verbs in declarative clauses, where {-en} infinitives would be favored over bare stem lexical verbs (Rice et al., 1997, p. 267). At the same time, SLI children are assumed to “know many of the properties of functional projections” and should thus be aware of the consequences of finiteness and Tns/Agr marking (Rice et al., 1997, p. 256). Hence, finite verbs are expected in V2/nonfinal position, whereas infinitival verbs are expected to appear clause-finally in main clauses. Moreover, if a verb is morphologically marked for finiteness, the person and number features on the verb must agree with the sentential subject.

Analyses on finiteness were carried out on lexical verbs and copular *sein*, while clauses containing modal verbs were only examined for sentence bracket (Satzklammer) effects, as they do not show overt 1st and 3rd PS SG inflectional marking. SVA was examined in a rather conservative manner and was limited to utterances with an overt

subject⁹² and one of the target verbs, in which “the verb is fully specified for Tns and Agr” (Hamann, 2015, p. 229; Rice et al., 1997), including utterances without a verbal complement. Moreover, only forms marked for 2nd PS SG {-st} and 3rd PS SG {-t} were considered for the SVA analyses due to the ambiguities inherent to the surface forms of the remaining verbal inflectional markings, e.g., 3rd PS PL {-n, -en} and due to the low frequency of some of the forms, e.g., 2nd PS plural marking {-t}.

As for finiteness markings, despite the significant difference between the SLI and TD groups with respect to the use of infinitival lexical verb forms and copula omissions (which were significantly more frequent in the SLI group), both groups, contra Clahsen et al. (1997), demonstrated high SVA⁹³ accuracy rates at both measuring points. TD children showed ceiling performance on 3rd PS SG {-t} at T1 and T2 and attained 80% accuracy levels at T1 and 100% at T2 on 2nd PS SG {-st}. Despite using fewer analyzable lexical verbs, the performance of the SLI group generally resembled that of the TD control group, with 88% accuracy scores on 3rd PS SG and 71% accuracy on 2nd PS SG at T1 and ceiling performance at T2 on both inflectional suffixes. Error patterns primarily involved substitutions of 2nd PS SG {-st} by {-t}, which, according to the authors, is very likely to result from articulatory difficulties. As to the copular verb *sein*, the prevalent error type in both groups was copula omission⁹⁴ (not substitution), with higher omission rates in the SLI group (SLI: 29% vs. TD: 17%) at T1 and almost target-like performance at T2 in both groups (<10% copula omission). Crucially, Rice et al. found that whenever copula *sein* was not omitted, it showed target consistent inflection in 97% of the cases by the SLI children, who also used several different forms of it.

Concerning finiteness and word order, although children with SLI were significantly more likely to use nonfinite forms of lexical verbs in declarative sentences than their younger MLU-matched typically-developing peers (SLI: 52% vs. TD: 8%) at T1 with a substantial improvement at T2 (SLI: 12% vs. TD: 8%), SLI children appeared to generally adhere to verb movement rules as finiteness and verb placement were highly correlated as is the case in the TD group. Results of the SLI group showed that most of

⁹² “Overt subjects ... establish an obligatory context for finiteness” (Rice et al., 1997, p. 272).

⁹³ Unlike Clahsen et al. (1997) and studies by Rothweiler and colleagues, who investigated agreement in contexts with fully specified DPs, Rice et al. (1997) examine agreement in contexts, in which the verb is fully specified for the features Tns and Agr.

⁹⁴ Note that within the EOI, auxiliary and copula omissions mark missing tense and agreement and not only agreement as in the ADH.

their non-finite verb forms were placed clause-finally⁹⁵ (T1: 43/45, T2: 29/29), a pattern that is compatible with the EOI account. As to finite verbs, whenever a finite verb was used, it mainly appeared in V2 position with high accuracy rates in both groups at T1 (SLI: 88% vs. TD: 94%) and T2 (SLI: 98% vs. TD 99%)⁹⁶. In sum, Rice et al. found that the SLI children almost always correctly inflected finite verb forms appearing in V2 and that root infinitives with *-en* forms were the predominant error type. Consequently, Rice et al. conclude that SVA is relatively spared in German children with SLI and in younger typically developing children and that the observed SLI patterns are compatible with the EOI account.

Hamann (2015) raises some important factors likely to be responsible for the discrepancy between Rice et al.'s and Clahsen et al.'s findings. Aside from differences in participant selection, coding, and scoring procedures (which highly depend on the underlying theoretical assumptions of German clause structure, UG's computational system and the models adopted for typical language development) and articulatory deficits that could obscure mastery of the SVA paradigm, Hamann (2015) lists dialectal variation as a crucial factor. Whereas the dialectal varieties spoken by the children investigated in Rice et al.'s study are carefully accounted for in scoring, little is known about the dialectal varieties spoken by the children examined in Clahsen et al. and subsequent studies. Given that Clahsen et al.'s data were collected around the Duesseldorf area, Hamann (2015) speculates that the children might have been exposed to colloquial varieties that feature aspects of "Platt Deutsch"- Low German, in which *-en* forms are typically elided, while preterite forms are more frequently used than in the Southern varieties spoken by the children in Rice et al.'s study leading to different results. Importantly, despite arguing that an agreement deficit is at the core of the grammatical

⁹⁵ Note that there were a few instances of separable prefix verbs (particle verbs) where the finite verb stem was adjoined to its particle in clause-final positions. In line with Grimm (1993) as well as Poeppel and Wexler (1993), Rice et al. argue that the latter forms are most likely attempted participial forms lacking the unstressed *ge-* due to articulation difficulties. Yet, they acknowledge that particle verbs are particularly challenging for children with SLI.

⁹⁶ Note that these figures are based on the adjusted count, in which the occurrence of *-en* forms in V2 position with plural subjects was tabulated. This was done because *{-en}* can serve as a finiteness marker when it co-occurs with a plural subject. Analyses using the conservative method, which considers all *-en* forms of lexical verbs non-finite even if they occur with plural subjects, yielded a rather unexpected pattern at T2. Based on the conservative count, more than 40% (20/49) of the non-finite verbs would occur in V2. However, the second analysis adjusting for plural subjects revealed that all of those instances were cases of plural subjects in combination with *-en* forms in V2, hence most likely to be finite (Rice et al., 1997, p. 279).

deficit in German SLI, Clahsen acknowledges that several linguistic phenomena that are vulnerable in children with SLI “fall outside what is covered by an impairment of agreement” (Clahsen & Dalakakis, 1999, p. 18). Examples of such phenomena are comprehension of reversible passives and reflexive anaphors in English (van der Lely, 1996; van der Lely & Stollwerck, 1997), appearance of non-finite verb forms in V2 in the spontaneous speech production of some German-speaking children with SLI as in Dieter’s case (Clahsen et al., 1997), and tense marking problems reported for English and Hebrew speaking children with SLI. As to the latter issue, Clahsen et al. (1997, p. 165), interestingly, mention that “sentences with non-finite verbs lack Comp and Tense and hence no verb-raising” implying that the occurrence of infinitives in clause-final position could arise due to the absence of the Tense feature, and “not [due] to an agreement deficit” (Hamann, 2015, p. 227).

In order to explain SLI phenomena that cannot be accounted for by the ADH, EOI (missing Agr or Tns features) or the Truncation Hypothesis (Rizzi, 1994), such as problems with building *wh*-questions and subordinated constructions in English (van der Lely, 1998) and Hebrew (Novogrodsky & Friedmann, 2006) as well as difficulties with structures involving contrastive focus and past perfect in German SLI, Hamann et al. (1998, 2001) proposed that children’s interim grammars adhere to a Minimal Default Grammar⁹⁷ (Roeper, 1996). Specifically, Hamann et al. argue for an impaired clause structure in German SLI that concerns “the IP/CP conflation”, due to which children “fail to project an extended structure [(i.e., a fully specified CP)] that unifies the CP shell with the propositional core of the clause” (Hamann et al., 1998, p. 228). Hence, the resulting projection of a target structure is underspecified in that it may contain the lexical features but only “a subset of the [relevant] formal features rather than full-fledged phrase structure categories with categorial and functional labels” (Hamann et al., 1998, p.204). In line with Travis (1991) and Zwart (1997), Hamann et al. (1998) emphasize the importance of differentiating between generalized and residual V2. In adult German grammar, (main) clause merging is the outcome of successive “head to head (V^0 -to- I^0 -

⁹⁷ The Minimal Default Grammar Hypothesis as established by Roeper (1996) and extended by Penner and Roeper (1998) postulates the following:

“All non-target (intermediate) forms would be reflections of a Minimal Default Grammar: The set of defaults is made available by Universal Grammar. All defaults will represent a more economical form in terms of ‘economy of derivation or economy of representation’” (Penner & Roeper, 1998). In other words, as long as the clause structure is not fully-specified, children would resort to a more economical form in terms of representation or derivation.

to $-C^0$) movement”, which occurs independent of “the categorical, semantic, and relational identity of the constituent in Spec CP” (Hamann et al., 1998, p. 199). The latter process is known as generalized V2 and is assumed to represent main clauses in adult German grammar. From this perspective, every declarative main clause constitutes a complex structure in German that involves a CP projection (Hamann, 2015, p. 236). On the other hand, children are assumed to initially resort to a somewhat selective V2 rule driven by Economy Principles, namely the *residual V2 rule*. The central assumption is that verb movement to C^0 is triggered only if the constituent raised to Spec CP fulfills a certain semantic function, i.e., has an operator status with the feature [+wh], or constitutes a topicalized or a focalized element (i.e., has a specific discourse function). Accordingly, object-initial clauses, constituent questions, and subordinate clauses unambiguously involve a CP shell. In contrast, subject-initial clauses do not necessarily involve a CP layer and do not trigger movement to Spec CP (in intermediate child grammar) since they are “neutral with respect to operatorhood or discourse representation” and can thus be accommodated within a projection lower than the CP, i.e., an IP or AgrP/TP as a maximal projection. Clauses with adverbs in first position have also been argued to occupy a position different from fronted objects (focalized or topicalized) for semantic and syntactic reasons. As a result, a CP layer is not necessarily projected in the case of adverb-initial clauses either, which, similar to subjects, can be accommodated within an intermediate projection between the VP and the CP, namely a head-initial AgrP/TP or an IP.

Although residual V2 does not conform to adult grammar, it represents “a logical option of the V2 rule” that is attested in other languages (Hamann et al., 1998, p. 200). As such, it can be viewed as the null hypothesis that needs to be verified and subsequently revised by the child through the course of language development. Children are said to initially adopt the more parsimonious residual V2 rule for subject and adverb initial clauses until they resolve the conflict between residual V2 rule (according to which subjects and adverbs occupy the specifier position in a head-initial IP or AgrP/TP) and generalized V2 (with subjects and adverbs occupying the specifier of CP regardless of their semantic function or operatorhood). Unlike their TD peers, German language-impaired children are argued to “stagnate before they can resolve the conflict between residual and generalized V2” (Hamann et al., 1998, p. 201). Consequently, they encounter persistent difficulties in capturing the “full feature specifications of C and its specifier(s)” which has not only to allow subjects to appear in this position but also adverbs and

objects, “and this is the defining feature of V2 languages” (Hamann, 2015, p. 232). Hamann (2015) stresses that, contrary to what is claimed by Rothweiler et al. (2012) and Ruberg et al. (2019), the Minimal Default Grammar does not predict the absence of a CP layer in children with SLI, which would predict them not to be able to produce questions and subordinated clauses at all. In contrast, the Minimal Default account assumes that children with SLI can project a CP, which, however, is not fully specified. As a result, they may demonstrate errors in structures with *wh*-movement, subordination, as well as generalized V2, and occasionally “leave finite verbs in clause-final position, i.e., in the IP” (Hamann, 2015, p. 232).

Hamann et al. (1998) examined spontaneous speech data of 50 monolingual German-speaking children with SLI (age range: 3;2-10;3 yrs.) stemming from six different corpora (own data from the Lindner corpus, 1995, Clahsen, CHILDES, 1988, Haffner, 1995, Hansen, 1994, and Penner, 1994⁹⁸). The study investigated different aspects of impaired clause structure in German, preferences for finite vs. non-finite verb forms, patterns of verb placement with respect to finiteness, as well as production of *wh*-questions and subordinate clauses, and compared the performance of their SLI participants with German TD data in published studies. Since Hamann et al.’s study was primarily concerned with the CP-domain, only declarative main clauses of the type (Object + V +XP) structures (i.e., sentences with topicalized objects) but not SVO or AdvVS structures were counted as instances of genuine V2. As previously mentioned, SVO and AdvVS structures resemble genuine V2 structures on the surface but are not necessarily generated by generalized V2 in child German; hence, they are considered ambiguous with respect to the highest projection that hosts them. Hamann et al. treated clauses of the latter type separately from genuine V2 under the category “pseudo-V2”. Cases with clitic pronouns⁹⁹ were also treated as pseudo-V2 given their ambiguous status between genuine V2 and Verbend (VE) structures (see Penner et al., 1992, for prosodic explanations). Unlike Rice et al. (1997), who counted particle verbs not separated from the finite verb in VE as attempted participles, thus non-finite, Hamann et al. (1998) treated

⁹⁸ Note that the children in Hamann et al.’s (1998) study were exposed to different dialectal varieties of German. Children from the Clahsen, Haffner and Hansen corpora were exposed to Northern varieties, the children in the Lindner corpus were exposed to Bavarian German as spoken in the city of Munich, while the children from Penner’s corpora were exposed to Swiss German. However, these dialectal differences are not of relevance for the linguistic phenomena analyzed in Hamann et al.’s study.

⁹⁹ In Southern German dialects, subject pronouns are clitics that function as “agreement markers on the verb, not full DP” (Hamann, 2015, p. 233).

such forms as unambiguously finite and clause-final. The following error types were considered as cases of target-inconsistent subordination: clauses lacking a complementizer (i.e., zero gap), a placeholder for the complementizer in the form of (pre)lexical placeholders, or subordinate clauses with infinitives [+C, Vinf] or a missing verb [+/-C, -V]. With respect to *wh*-questions, target-inconsistent responses included the following: missing *wh*-pronoun, a placeholder for the *wh*-pronoun, clause-final *wh*-questions (in unambiguously non-echo questions), questions with missing or infinitival verbs, and dropped subjects, see Hamann (1998, p. 212f) for examples.

Contrary to Rice et al. (1997) and Clahsen (1991b), Hamann et al. (1998) found that children with SLI used “more finite than nonfinite verb forms (57% vs. 36%)”, and as a group showed a preference for clause-final finite verbs (44%) over genuine V2 (fronted objected, 3%) or subject-initial (27%) verb-second patterns in declarative main clauses suggesting that the investigated children have not acquired the generalized V2-rule (Hamann, 1998, p. 216). Despite slight differences between the corpora¹⁰⁰, the pattern of finite verbs occurring in VE was attested in all of the examined corpora affecting more than 50% of all unambiguously finite verbs in some of them (particularly Haffner, 1995, and Penner, 1994, 1998). Nevertheless, the general trend observed in the compiled data is that SLI children overuse VE patterns in comparison to genuine or subject-initial pseudo-V2 constructions (Hamann et al., 1998, p. 218); hence, the close link between finiteness and verb placement in C does not seem to hold for monolingual children with SLI. This pattern greatly diverges from that observed in TD children, who are reported to place finite verbs in VE to a much lesser extent (10-20%), see Clahsen et al. (1996a); Penner (1990). Consequently, Hamann (2015) argues that the impairment is “much more than a mere delay” (p. 235). Further support for deviance comes from the analyses of *wh*-questions in Hamann et al.’s (1998) study. Not only did the SLI children show high rates of target-inconsistent questions in 76% of the cases, but they also demonstrated a unique error pattern, namely *wh*-questions with infinitives, i.e., the occurrence of a *wh*-pronoun together with an infinitive instead of a finite verb, which is not attested in typical development (Clahsen et al., 1996a). The latter pattern was the prevailing error type along with *wh*-questions, in which the finite verb appeared in final position. Moreover, children with SLI experienced exceptional difficulties producing target-like subordinated

¹⁰⁰ The children in the Clahsen corpus showed a preference for pseudo-V2 of the type Subj +V+ XP, whereas the children from the Lindner corpus showed no particular preference, which, according to Hamann et al. (1998), might reflect data sampling effects.

constructions (less than 20% target-consistent). Qualitative analysis revealed a trend¹⁰¹ of producing subordinate clauses with a missing complementizer and subordinate clauses in which an overt complementizer occurs together with an infinitive. Based on their results, Hamann et al. (1998) entertain the possibility that a broader impairment could be the source of SVA and verb placement problems observed in German SLI and other V2 languages like Dutch (de Jong, 2004). Such an impairment is linked to the notion of “computational complexity” and affects the grammatical computation of complex derivations (Jakubowicz, 2004, 2011; van der Lely, 2005a; see also Hamann, 2006; Hamann et al., 2007). Given that German is a genuine V2 language, every German matrix clause is inherently complex as it features a CP projection in adult grammar, in which V2 is generated via consecutive V to I⁰ to C⁰ movement. In this sense, children’s errors with agreement and finiteness markings could reflect children’s problems with computational complexity; see 2.1.2.2 for a detailed description.

As to the acquisition of verbal inflectional morphology and verb placement in child bilinguals acquiring German, several studies based on spontaneous speech data and elicited production documented close parallels between BiTD children with an AoO to German between ages 2 and 4 and MoTD German-speaking children (Chilla, 2008; Lemmer, 2018; Prévost, 2003, Rothweiler, 2006; Schulz & Schwarze, 2017; Schulz et al., 2017; Tracy & Thoma, 2009), where mastery of the SVA inflectional paradigm and correct V2 placement is achieved within 6-18 months of exposure to the L2. Before that, bilingual children are reported to go through an optional infinitive-like stage, fluctuating between target-consistent finite verbs in V2 and infinitives in clause-final positions (Rothweiler, 2006; Thoma & Tracy, 2006). During this phase, BiTD children rarely produce substitution errors or infinitives in V2, in contrast to MoSLI children (Rothweiler, 2006), with a tendency to overuse bare forms in V2 (Chilla, 2008; Prévost, 2003; Rothweiler et al., 2012). Bare forms in V2 have also been observed in productions of bilingual children acquiring L2 Dutch (Blom & Baayen, 2013) and are also reported for English (though a non-V2 language) as shown by Haznedar (1997a, 1997b) and Haznedar and Schwartz (1997). If one adopts Clahsen et al.’s (1997) as well as Rothweiler and collaborators’ stance, who consider bare forms potentially non-finite, one could argue that eL2 children have SVA problems, which, at least on the surface, would be interpreted

¹⁰¹ Given the limited database and different distributions with the compiled corpora, the latter results should best be interpreted as trends that need to be verified in controlled experimental tasks (Hamann et al., 1998, p. 222).

as an overlap with patterns attested in MoSLI children (Chilla, 2008; Kroffke & Rothweiler, 2006). However, adopting the *Missing Surface Inflection Hypothesis* (MSIH), Prévost (2003) argued in his case study with an eL2 TD child that bare stems occurring in V2 (a finite position) are covertly finite, i.e., they are specified for finiteness [+finite] but underspecified for the features ‘number’ and ‘person’. Wojtecka et al. (2013) addressed the issue of bare forms in their study with 25 eL2 (AoO: 3;0) BiTD children with diverse first languages. The children were assessed twice: once after ten months (around age 3;9) of exposure to the L2 and once after 19 months of exposure (around age 4;8). Their results showed that at T1, 76% of the verbs appearing in V2 were correctly inflected for SVA. Non-target consistent verb forms in V2 were mainly bare forms (14%), with true *-en* infinitives occurring only rarely in V2 (7%). Moreover, errors in clause-final positions were predominantly unambiguous *-en* infinitives (83%), whereas bare forms rarely appeared in VE (11%), and there were no instances of finite verb forms in clause-final positions. At T2, eL2 children showed correct finiteness marking and verb placement in 97% of the cases, suggesting that they have mastered finiteness and verb placement. The authors argue that bare forms “function as a default and reflect eL2 children’s difficulty with the realization of the correct overt morphological suffix”, as proposed by the MSH (Schwarze et al., 2015, p. 434). Accordingly, bare forms are likely finite because they rarely appear clause-finally and predominantly appear in V2 position, which is typically reserved for finite verb forms. In line with Prévost (2003), Wojtecka et al. (2013) conclude that eL2 BiTD children resemble their MoTD peers in being aware of the relationship between V2 and finiteness (see also Schulz & Schwarze, 2017). Hence, Schulz et al. (2017) and Schulz and Grimm (2019) assume that SVA can be used as a clinical marker for eL2 children with limited exposure to L2. Late successive children (AoO > 6;0), on the other hand, do not appear to adhere to the ban of finite verbs in V2 and produce non-finite verb forms in V2 (Chilla, 2008; see also Sopata, 2011).

As to BiSLI children, Schulz and Schwarze (2017) investigated potential links between finiteness and verb placement in BiSLI children using data elicited via the LiSe-DaZ test (Schulz & Tracy, 2011). They compared the performance of 11 eL2 older BiSLI children (mean age at T1: 7;1) to 22 younger eL2 BiTD children (mean age at T1: 3;7) at four testing rounds with a 6-month interval at the first three rounds and a 12-month interval at T4. In this study, Schulz and Schwarze (2017) differentiated between bare forms and infinitives and between V2 and verb-final forms. Their results showed that across all test rounds, eL2 children with SLI demonstrated high rates of correct SVA in

V2 with 83.3% at T1 and ceiling performance at T4 (mean age: 8;11). Substitution errors were practically absent, and *-en* infinitives in V2 occurred only in (1.08)% of the data. On the other hand, bare forms in V2, with an overall rate of 11.4%, constituted the most frequent non-target-like form. Contrary to their BiTD peers, BiSLI children produced some bare forms (8.1%) and finite verb forms in clause-final positions (13.5%). Subordinate clauses with an overt complementizer showed a distinct pattern, with a 94% rate of target-consistent inflection in BiTDs and 88% in BiSLIs. Since finite verb forms and bare stems, which Schulz and Schwarze view as finite, mainly occurred in V2, and *en*-infinitives occurred more frequently clause-finally than in V2, Schulz and Schwarze (2017) conclude that the “ban on non-finite verbs in V2 is [inviolable] ... in eL2 TD and in [7-9-year-old eL2 BiSLI children]..., indicating that SLI is a matter of severe delay... that concerns the overt realization of the phi-features person and number on the verb (p. 51ff). Nevertheless, given that substitution errors were not attested in their data, Schulz and Schwarze argue that their data do not support Clahsen et al.’s (1997) ADH.

It should, however, be noted that the children in Schulz and Schwarze’s (2017) study were older than 8;0, showed mastery of the German verbal inflectional, and would therefore have been underdiagnosed for SLI based on this clinical marker. This implies that problems with finiteness markings might not be suitable for identifying SLI in older bilingual children. This is in line with the findings of Scherger’s (2019b) study, which evaluated the diagnostic potential of a LITMUS-SVA elicitation task (de Jong, 2015) with a sample of 22 eL2 BiSLI and BiTD children aged six to eight years. Scherger’s results showed that despite the significantly lower performance of the BiSLI children relative to their age-matched BiTD peers on SVA, some BiSLI children appeared to have already mastered the SVA paradigm, yielding an overall diagnostic accuracy of only 82%. Consequently, Scherger recommends complementing SVA with other tasks, as it appears to be not sensitive enough in the case of older bilingual children.

In sum, regardless of the theoretical assumptions concerning the core deficits in the domain of verbal inflectional morphology, be it tense, agreement, or finiteness in general, it is pretty apparent that both monolingual and bilingual children with SLI have deficits in verbal inflection. However, given the surface overlap between BiTD and MoSLI children, especially in the form of bare stems, and the limited sensitivity of this marker to the impairment in older children, sole reliance on this marker for the identification of language impairment is not possible, and the test should be

complemented with tools that utilize markers that show unambiguously robust acquisition in bilingual children with no (surface) overlap with MoSLI errors.

Complex syntax

A large body of cross-linguistic research has shown that children with SLI demonstrate marked difficulties in the comprehension and production of computationally complex constructions, such as relative clauses, object *wh*-questions, and embedded constructions, e.g., finite complement clauses (Adani et al., 2014, 2016; Friedmann & Novogrodsky, 2004, 2011; Hamann et al., 2007, 2017; Hamann & Tuller, 2014; Jensen De López et al., 2014; Owen & Leonard, 2006; Paradis et al., 2017b; Roesch & Chondrogianni, 2015; Scheidnes & Tuller, 2014, 2019; van der Lely, 1998; van der Lely et al., 2011). A prominent hypothesis that builds upon the notion of “computational complexity” is the Computational Grammatical Complexity Deficit hypothesis (CGC; Marshall & van der Lely, 2006; van der Lely, 2004, 2005a). The CGC is a development of the Representational Deficit for Dependent Relations (RDDR) account (van der Lely, 1997a, 1998). It postulates that children with grammatical SLI have a domain-specific deficit in the computational system responsible for the construction of hierarchically complex constructions involving non-local dependencies, e.g., filler-gap dependencies as in *wh*-questions and relative clauses. In line with this, constructions featuring aspects of computational complexity (Jakubowicz, 2004, 2011; van der Lely, 2004, 2005a) and combinations thereof, such as embedding, A/A' movement, and intervention, are assumed to be particularly taxing for language-impaired children given their limited WM resources compared to their typically developing peers (Adani, 2011; Arosio et al., 2011; Delage & Frauenfelder, 2020; Grillo, 2008; Montgomery et al., 2010). Children with SLI have been found to exhibit significantly higher error rates than their typically developing peers or avoid complex constructions altogether, both in spontaneous and elicited production (Hamann et al., 2007, 2017; Hamann & Tuller, 2014; Marinellie, 2004; Owen & Leonard, 2006; Rakhlin et al., 2016; Scheidnes & Tuller, 2014, 2019). For an elaborate description of the factors contributing to computational complexity and an in-depth discussion of some of the studies mentioned above, the reader is referred to section 2.1.2.2

Studies on the comprehension of complex structures in morphologically-rich languages have revealed that, in comparison to their TD peers, children with SLI are less efficient in using morphological markings, such as number marking on the verb and case,

as disambiguating cues in complex non-canonical structures, in which thematic role assignment does not rely on linear word order (e.g., Roesch & Chondrogianni, 2015; Stavrakaki, 2001; Stavrakaki et al., 2015; Stegenwallner-Schütz & Adani, 2021). For example, Stavrakaki et al. (2015) found that, unlike their younger TD controls, monolingual Greek-speaking children with SLI (ages 5 to 8) did not profit from the presence of case marking for the interpretation of object relatives. Similarly, in their study on the comprehension of object questions in German, Roesch and Chondrogianni (2015) observed that a mismatch in the case feature between the *wh*-element and the subject DP in referential object-which questions enhanced comprehension in the TD group but not in the SLI group. Stegenwallner-Schütz and Adani (2021) investigated the role of number features in the comprehension of SVO and topicalized OVS sentences in 45 German children with and without SLI (age range: 5;0-10;11, mean age: 6;9). In line with previous studies documenting poorer comprehension of OVS than SVO structures up to age seven in MoTD German-speaking children (e.g., Biran & Ruigendijk, 2015; Ruigendijk & Friedmann, 2017), OVS structures were comprehended less well than SVO structures not only by SLI but also by TD children, with larger effect sizes for the SLI group. Interestingly, number mismatch in the form of a singular object and plural subject NP facilitated comprehension of OVS structures in both SLI and TD groups, which were better comprehended than OVS with two singular NPs (see Adani et al., 2014, for similar findings on relative clauses). The authors argue that the facilitative effect of number mismatch arises due to “the reduction of feature similarity between the subject and the object” (Stegenwallner-Schütz & Adani, 2021, p. 7; see also Adani et al., 2017; Belletti et al., 2012), which would reduce their computational complexity and the associated processing load. Stegenwallner-Schütz & Adani (2021) further argue that the facilitative role of number agreement goes against the ADH, which assumes a specific impairment in the number features of the verb. According to the authors, the fact that children with SLI can exploit number agreement between the subject and the verb, though less effectively and to a much lesser extent than TD children, suggests that the “computation of agreement relations follows the same feature-checking mechanisms as in typical development” (Stegenwallner-Schütz & Adani, 2021, p. 9). Concerning error patterns, both groups mainly resorted to thematic role assignment via canonical order opting for SVO interpretations. However, this pattern occurred significantly more frequently in the SLI group, corroborating previous findings on German children with SLI, who show poor comprehension of various computationally complex constructions (e.g., object *wh*-

questions: Roesch & Chondrogianni, 2015; object relative clauses: Adani et al., 2016, 2017).

In contrast to children with SLI, computationally complex constructions represent a relative area of strength for typically developing bilingual children, who can draw on their intact language faculty and working memory, and are thus expected to be capable of dealing with different operations such as recursion, movement, and embedding, given sufficient exposure to the L2 (Hamann et al., 2017). As shown in the studies synthesized in Section 2.1.2.2, bilingual children use complex constructions from early on, and compared with vocabulary and inflectional morphology, need significantly less exposure to the L2 to pattern like their monolingual age-matched peers. Taken together, the acquisition of complex syntactic constructions represents a relative area of strength in typically developing bilingual children, as it appears to proceed at a faster rate than other domains like lexicon and inflectional morphology. Hence, tasks featuring computationally complex constructions offer promising tools for establishing typical language development in bilingual children.

2.2.1.2 Phonological abilities of children with SLI

Albeit most deficient, morphosyntactic deficits in children with SLI often go hand in hand with phonological problems. In fact, most early referral cases to SLP services feature SLI children with concomitant phonological problems, as those are the easiest to identify. A large body of cross-linguistic research has demonstrated that language-impaired children also exhibit remarkable deficits in the domain of phonology, e.g., Arabic (Shaalán, 2020), English (Rescorla & Ratner, 1996; Roberts et al., 1998), French (Ferré et al., 2012), German (Fikkert & Penner, 1998; Fox-Boyer, 2016), Swedish (Nettelbladt, 1992), and Italian (Orsolini et al., 2001). Children with SLI lag behind their age-matched TD peers in the acquisition of consonants (Roberts et al., 1998), experience more difficulty forming accurate phonological representations (Alt & Plante, 2006; Claessen & Leitão, 2012; Maillart et al., 2004; McKean et al., 2013) and perform more poorly than TD children on measures of phonological processing¹⁰² (Claessen et al., 2013; Rispens & Baker, 2012),

¹⁰² According to Wagner and Torgesen (1987), phonological processing in its broad sense consists of three separate yet interconnected components:

such as nonword repetition (Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990), rapid automatized naming and phonological awareness tasks (Thatcher, 2010).

Inaccuracies in nonword repetition have been taken as evidence that limitations in phonological memory capacity (PSTM account: Botting & Conti-Ramsden, 2001; Gathercole, 2006a; Gathercole & Baddeley, 1990), in particular in the phonological loop component (Baddeley, 2003), are responsible for language deficits in children with SLI. The latter WM deficits have been argued to affect “the creation of a phonological entry within the long-term phonological store” (Archibald & Gathercole, 2006, p. 687). As such, poor performance on nonword repetition tasks has been proposed to be a reliable phenotypical marker for SLI in many languages (Bishop et al., 1996; Conti-Ramsden et al., 2001a; see Chiat, 2015, for an overview). Moreover, some studies showed that the difficulties children with SLI encounter with repeating nonwords that are two to four syllables long are likely to have a hereditary component (Bishop et al., 1996; Bishop et al., 2006; Kalnak et al., 2014). On the other hand, many studies have demonstrated that limitations in the ability to temporarily hold phonological material in working memory are not the sole source for poor performance on nonword repetition tasks and that performance is influenced by the phonological characteristics of the stimuli, such as word-likeness, phonotactic probability, neighborhood density, and structural complexity. For instance, several studies reported a facilitative effect of word-likeness on nonword repetition and that the repetition of nonwords of high word-likeness is correlated with lexical abilities (e.g., Engel de Abreu et al., 2013; Gathercole, 2006b; Hoff et al., 2008; Metsala, 1999).

-
- i. Phonological awareness (PA): “the awareness of the sound structure, or phonological structure, of a spoken word” (Gillon, 2004, p. 2), which is greatly implicated in the development of reading and literacy skills (Melby-Lervåg, et al., 2012). PA is commonly assessed via tasks that require manipulation of sound structure, e.g., addition or elision of segments or syllables to words or nonwords.
 - ii. Phonological short-term memory (PSTM), i.e. phonetic coding in working memory: temporary storage of phonological information in working memory that enables manipulation of such material. PSTM is usually measured by tasks tapping into the central executive and the phonological loop, e.g. nonword repetition and serial digit recall tasks.
 - iii. Phonological retrieval from long-term memory, i.e., “phonological recoding in lexical access” (Claessen et al., 2013, p. 472). This component is typically assessed via rapid automatized naming tasks, in which the participant is presented with an array of object pictures, graphemes or numerals and is required to name each of them as quickly as possible. Poor performance on such tasks is argued to be associated with underspecified phonological representations hindering accessibility (Snowling, 2000); however, oromotor weaknesses could contribute to slow performance as well (Wolf & Bowers, 1999).

With regard to phonological complexity, children with SLI are reported to be sensitive to complexity features such as the presence of consonant clusters (Bortolini & Leonard, 2000; Gallon et al., 2007; Ferré et al., 2012; Marshall et al., 2002; Marshall & van der Lely, 2009; Shaalan, 2020), coda position (Tamburelli & Jones, 2013), syllabic position in the foot (Bortolini & Leonard, 2000), and non-canonical prosodic word structure (Marshall et al., 2002). Gallon et al. (2007) and Marshall et al. (2002) found difficulties with nonwords featuring aspects of phonological complexity to persist into later childhood and even adolescence. In particular, language-impaired individuals appear to struggle with the “constraints that govern the association between consonants and syllable” (Ferré, 2012, p. 19), i.e., they encounter difficulties linking segments to their syllabic positions. As coping strategies, complex phonological structures are often simplified or even avoided in the form of consonant cluster reduction, resyllabification, or deletion of syllables, especially in unstressed positions (Bortolini & Leonard, 2000; Fox-Boyer, 2016; Marshall et al., 2003; Orsolini et al., 2001). The latter findings have been taken as evidence showing that nonword repetition is not merely a measure of phonological short-term memory but that it also taps into “the storage of, and access to, phonological representations” (Claessen et al., 2013, p. 472; Dollaghan et al., 1995; Rispens & Baker, 2012).

Studies looking into the nature of phonological representations¹⁰³ of words in SLI deliver further evidence that children with SLI struggle not only with holding phonological information in working memory but also have difficulties forming precise and finely detailed phonological representations (Edwards & Lahey, 1998). Phonological representations in children with SLI are argued to be rather underspecified and less stable than those of their typically-developing peers (Leonard, 2014a, p. 67; Maillart et al., 2004). For example, in order to assess phonological representations while minimizing

¹⁰³ The term phonological representation refers to the sound-based codes stored in long-term memory for each lexical item (Anthony et al., 2010; Gillon, 2002). In terms of quality, phonological representations can be either holistic, i.e., words are stored and produced as non-segmentable whole (single) units, or segmental, i.e., words are made up of subunits, e.g., phonemes and syllables, which are subject to conscious manipulation (Sutherland & Gillon, 2005). The quality of these representations is said to gradually develop by age starting by holistic representations in early years (Maillart et al., 2004; Snowling & Hulme, 1994) and evolving to fine-grained segmented units as children’s language experience and phonological awareness increase, see the “segmentation hypothesis” by Fowler (1991) and the “lexical restructuring hypothesis” by Metsala and Walley (1998). According to Fowler (1991), precise and well-specified phonological representations are crucial for efficient lexical retrieval and differentiating between similar sounding words.

output processing load, Claessen et al. (2013) employed the Quality of Phonological Representations task (Claessen et al., 2009). In this receptive auditory lexical discrimination task, the child must judge the accuracy of a polysyllabic word, e.g., “Is this a rhinoterous?” The study's results showed that children with SLI (age range: 6;10-8;0) were more likely to regard an inaccurate production of a multi-syllabic word as correct compared to their age-matched and younger language-matched typically-developing peers. The authors conclude that the quality of the underlying phonological representations is poorer and less detailed in the SLI group, especially for longer and more complex words. Another phonological priming study by Gray et al. (2012) investigated phonological word-form representations in preschool children with SLI using cued shadowing. Children listened to word pairs that included different types of primes (overlapping onsets or rhymes, identical or unrelated words) and were requested to repeat the second item as quickly as possible. Unlike the age and vocabulary-matched typically developing control groups, no inhibitory or priming effects were observed for the group with SLI. The authors argue that the absence of inhibition effects in the language-impaired group is attributable either to their inability to process the prime or target items or to “their phonological representations [that are] not detailed enough to prime words in their lexicon” (p.32), as low-quality phonological representations are proposed to be more difficult to segment and manipulate than high-quality ones (Elbro et al., 1998). Degraded phonological representations can further impede encoding, storage, and retrieval processes (Mainela-Arnold & Evans, 2005). Taken together, it can be assumed that long-term linguistic knowledge is implicated in phonological processing in line with Baddeley’s (2003) WM multi-component model and that the phonological deficits in SLI are not solely caused by limitations in the ability to hold information in working memory.

Furthermore, phonological limitations are assumed to impede subsequent lexical and (morpho)syntactic development in affected individuals as weaknesses in phonological processing are argued to “compromise the ability to form precise phonological categories, and consequently” (Coady & Evans, 2008, p. 14) disrupt the mapping¹⁰⁴ processes that are crucial for establishing structures at the word and sentence levels (Chiat, 2001; Marshall & van der Lely, 2006; Montgomery, 1995; Schwartz, 2009;

¹⁰⁴ Mapping theory postulates that, in order to acquire language, children have to segment the aurally incoming speech stream into individual words and morphemes, and map phonological form onto meaning, i.e., assign meaning to the segmented entities with the support of context while utilizing different cues, such as prosodic and semantic/lexical ones (Chiat, 2001).

Song et al., 2009; Tomas et al., 2015). Such weaknesses have also been found to be associated with subsequent difficulties in developing literacy and reading skills (Claessen & Leitaõ, 2012).

2.2.1.3 Lexical-semantic abilities in children with SLI

Despite being less severe than morphosyntactic weaknesses, especially in pre-school years, delays and deficits in lexical development are among the early indicators of SLI (Leonard, 2014a). From the outset of the language acquisition process, children with SLI lag behind their typically developing chronological-age peers in their acquisition of the lexicon (Dale et al., 2003; Leonard & Deevy, 2004). A comparative study by Trauner et al. (1995, 2000) using a parent report measure showed that children with SLI produced their first words almost at the age of 2;0 (23 months) as opposed to their typically developing peers, who reached this milestone around their first birthday. Not only do children with SLI produce their first words at a markedly older age, but they also demonstrate a slower rate of vocabulary growth and perform below age levels in terms of lexical diversity and vocabulary size both in production and comprehension (Kan & Windsor, 2010; Leonard et al., 1999; Thal et al., 1999; Watkins et al., 1995). Children with SLI are further reported to have a “relatively shallow understanding of the words they do use” (Leonard & Deevy, 2020, p. 3253; McGregor et al., 2013) and need more exposures to novel lexical items in order to achieve a learning level similar to their typically-developing peers (Alt et al., 2004; Gray et al., 2014; Rice et al., 1994). Over time, they tend to fall behind their typically developing peers (Rice & Hoffmann, 2015).

By school age, children with SLI are commonly reported to show poor word-learning¹⁰⁵ (Alt & Spaulding, 2011; Jackson et al., 2016; Storkel et al., 2017; see Kan & Windsor, 2010 for a meta-analysis) and deficient lexical retrieval abilities, which are not only delayed but are assumed to be qualitatively different from those of children with typical language development (Novogrodsky & Kreiser, 2015). These deficits are found

¹⁰⁵ Different accounts have been proposed to explain the complex process of word learning. In their influential paper, Carey and Bartlett (1978) first introduced the process of “fast-mapping”, that is the ability to form an initial association between a label and its referent after only a few exposures to the word (Chiat, 2001; Rice et al., 1992), i.e., a tentative map of the phonological and semantic representation (Leonard, 2014a, p. 55). They further argued that “fast mapping is only a small fraction of the total information that will constitute a full learning of the words” (Carey & Barlett, 1978, p. 2). Upon repeated encounters with the novel word, the latter process is followed by a “slow-mapping” phase, during which the initial phonological and semantic representations are gradually refined and further specified across contexts. This results in strengthening the connections between these representations (Carey, 2010).

under different labels in the literature, e.g., word-finding difficulties (Dockrell et al., 1998; Messer & Dockrell, 2013; Novogrodsky, 2015), Lexical-SLI (Friedmann & Novogrodsky, 2011), and Naming Difficulties (Faust et al., 1997).

Several explanations have been put forward to account for word learning and lexical retrieval difficulties in children with SLI. Two crucial distinctions relevant to word learning are “between word form and meaning and between encoding and long-term retention” (Leonard & Deevy, 2020, p. 3253). The word form denotes the phonetic form (phonological representation) of a word, e.g., /tɑɪɡə/ <tiger>, that is separate from its meaning (a wild feline carnivore animal). Despite having difficulties learning word meaning, individuals with SLI appear to struggle particularly with word forms (Gray, 2004; McGregor et al., 2013).

In order to retain a word over the long term and put it in a storable form, it has to be encoded in a precise and detailed manner (Alt & Spaulding, 2011), i.e., an accurate phonological and semantic representation has to be created that would allow its proper integration into the semantic and associative network. It is precisely the latter mechanism, namely initial encoding, which has been argued to be particularly compromised in children with SLI, especially for longer words (Alt, 2011; McGregor et al., 2017). This, in turn, impedes word retrieval in children with SLI, “not because the retrieval mechanism is selectively disrupted” (Leonard, 2014a, p. 59) but because retrieval is rather inefficient due to their underspecified lexical representations. This is reflected in knowing fewer details about words and having less elaborated lexical networks leading to weaker links between lexical items, see the “storage-elaboration hypothesis” (Kail et al., 1984; Kail & Leonard, 1986; Rothweiler, 1999); see also McGregor et al., 2013; Sheng & McGregor, 2010). This is in line with Rothweiler (1999), who found fast mapping not to be impaired in SLI. Rothweiler (1999) argues that word-learning difficulties arise either due to problems with initial encoding, which hinder the effective integration of new words into the associative network, or due to problems with “storage and access [as a result of] insufficient organization of the network” (Hamann, 2015, p. 239; see also Gathercole & Baddeley, 1993). A further factor assumed to contribute to lexical retrieval difficulties is the poor quality of phonological representations in children with SLI, which negatively affects “encoding, storage, and retrieval of [lexical] items” (Montgomery et al., 2019, p. 3821; see also Mainela-Arnold & Evans, 2005).

Verb learning in children with SLI

With respect to the development of lexical categories, it has been claimed that children with SLI “follow a different strategy in building their verb lexicon and show deviant development in the area of verb acquisition” (Hamann, 2015, p. 24; Penner et al., 2003; Schulz et al., 2001). During the single-word stage, children with SLI appear to resemble younger typically developing children with respect to the types of words and lexical errors they produce. Both groups acquire nouns before verbs, where names of objects, substances, and animals represent more than 50% of the used words, and tend to overextend words to further referents on the grounds of perceptual and functional similarity (Leonard & Deevy, 2004; Leonard et al., 1982; Weiss et al., 1983). In the multiword-utterance phase, qualitative differences begin to emerge with verbs becoming disproportionately challenging for children with SLI (Leonard & Deevy, 2004; Oetting et al., 1995) in both production and comprehension, especially in the case of low-frequency verbs (Loeb et al., 1996) or certain types of verbs like resultative¹⁰⁶ (telic) verbs (see Ingham et al., 1998 for English, Penner et al., 2003; Schulz, 2018 for German, and van Hout, 2000, 2018 for Dutch). It should be noted, however, that even for children with typical language development, the acquisition of verbs is relatively more difficult than the acquisition of object-referring nouns (see Guasti, 2002, p. 80ff; Hamann, 2015, p.238ff; and Schulz, 2018, p. 124 for discussions).

Several explanations have been proposed to account for the exceptional difficulties children with SLI encounter in establishing verb meaning. Under the assumption that impaired phonological processing is the primary culprit in language deficits in SLI, Chiat (2001) proposes that prosodic aspects pertaining to verb phonology contribute (along with other factors such as semantic and morphosyntactic ones) to the difficulties children with SLI¹⁰⁷ encounter with the acquisition of verbs. Chiat (2001) notes that verbs are seldom presented in isolation in contrast to nouns, which children frequently encounter as individual word forms. Verbs tend to be mainly embedded in

¹⁰⁶ A resultative verb phrase is a VP that contains an oblique constituent (PP or particle) that is not lexically selected (not obligated) by the verb but functions as an argument, e.g., the prepositional phrase *under the bed* in the sentence *the child rolled the football under the bed* (Leonard & Deevy, 2004, p. 225). It thus expresses “both an action and the state of affairs resulting from the action” (Ingham, 1998, p. 90). Note that some Languages like French and Japanese do not allow such constructions (Snyder, 1995).

¹⁰⁷ Difficulties with verb phonology are also postulated to contribute to verb difficulties encountered by TD children; however, children with SLI are expected to be disproportionately affected by such aspects given their limitations with complex phonological processing (Chiat, 2001).

sentences, are shorter in duration, and tend to bear non-canonical stress patterns. This makes them phonologically less salient than the neighboring nouns and harder to segment from continuous speech leading to increased perception difficulties (Black & Chiat, 2003).

Aside from phonological processing accounts, it has been argued that syntactic deficits underlie the difficulties children with SLI face in deducing verb meaning from context. In particular, it is the mechanism of syntactic bootstrapping¹⁰⁸ that has been invoked as a crucial process for extracting verb meaning from the syntactic frame of the sentence (Froud & van der Lely, 2008; Oetting, 1999; van der Lely, 1994), i.e., the argument structure in which the verb appears (see Leonard, 2014a, p. 70-74 for an overview). While such interpretative mechanisms have been found to be intact in young typically developing children (Gleitmann, 1990; Naigles, 1990), children with SLI appear to have difficulties exploiting them (van der Lely, 1994) due to their grammatical limitations in general as put by Rice (1991, p. 455): “The end result would be the opposite of bootstrapping. Instead of using one area of language to build another, children with SLI would be left without a solid strap to hang onto”. In the same vein, Ingham et al. (1998) found that English-speaking children with SLI (mean age: 6;0 yrs.) encountered more difficulties with complex resultative VPs, e.g., knock the ball (off the table), than their age-matched and younger vocabulary-matched typically developing peers. They attributed the vulnerability of complex resultative predication in SLI to “deficits with structurally determined [nonbinary dependent] relations between form and meaning”, i.e., to “deficits in the interface between syntax and the verb lexicon” (Ingham et al., 1998, p. 107).

¹⁰⁸ Note that Pinker (1994, p. 387) prefers using the term ‘syntactic cueing of word meaning’ over ‘syntactic bootstrapping’ to refer to the mechanism, by which verb meanings are learned via their syntactic context.

On the other hand, studies focusing on the acquisition of telicity¹⁰⁹ in (German) speaking children with SLI (e.g., Penner et al., 2003; Schulz, 2010; Schulz & Wittek, 2003; see Schulz, 2018 for an overview and a discussion) point to the existence of other strategies not relying on argument structure that could be involved in establishing verb meaning. The latter are more compatible with the semantic bootstrapping hypothesis in terms of Pinker (1994), which assumes that “the semantics of at least *some* verbs [are] acquired without relying on syntax” (p. 386); see also Hamann (2015) for a discussion. Schulz and Wittek (2003) investigated the comprehension of telicity in 5-year-old typically developing children and children with SLI. Their results showed that while the language-impaired group performed significantly worse than the TD group on inherently telic particle verbs, they demonstrated adult-like performance on atelic verbs (> 90% correct). These findings suggest that children with SLI are less sensitive to “the event component ‘endstate’, involved in telicity” (Schulz, 2018, p. 124), which appears to be present at a relatively early age in TD children (see Schulz et al., 2001; Schulz & Penner, 2002). Another study by Penner et al. (2003) looked into the production and comprehension of telic verbs and included older children with SLI (age range: 5;0-8;7) in addition to a younger SLI group (age range: 2;11-4;10). A clear difference emerged between the TD and SLI groups in that TD children (even the youngest ones) primarily opted for particles marking endstates in production and exhibited high rates of correct endstate interpretations in comprehension. In contrast, even older children with SLI demonstrated a bias for process particles in production, where a substantial proportion of them accepted a telic reading in atelic contexts in the comprehension tasks. In light of these results and findings of related studies, Penner et al. (2003) proposed the *Event Structural Bootstrapping* account as a strategy for building the verb lexicon, according to which TD children “initially focus on the verb’s event structure rather than on its core

¹⁰⁹ Telic verbs “designate so-called endstate-oriented transitions” (Schulz, 2018, p. 125) as opposed to process-oriented verbs like *sweep*. In Germanic languages like Dutch, English and German, “(a)telicity of the verb is generally determined by its lexical semantics or via the interaction with the morpho-syntactic context the verb appears in” (Schulz, 2018; see also van Hout, 2008). In German, telicity can be semantically triggered via entailment, either inherently, where it forms part of the verb meaning as in the verbs *finden* ‘find’ and *aufmachen* ‘open’, or compositionally through morphological marking via telic verb particles (also known as resultative verb particles), e. g. *auf-* (*aufessen* ‘eat up’ vs. *essen* ‘eat’). The latter particles are viewed as strong telicity markers since they obligatorily encode event completion when added to an atelic process verb. Telicity can also be pragmatic, where an event completion is not entailed but rather arises through implicature of the culmination point. This is achieved compositionally by adding certain quantized NPs such as specific definite NPs (weak telicity markers, see Schulz, 2018 and references therein for a detailed description, see also Hamann, 2015 for a discussion).

meaning or on its argument selection” (Schulz, 2018, p.130; see also Schulz et al., 2001 & Schulz et al., 2002). Under this account, TD children start acquiring verb semantics with a strong preference¹¹⁰ for *Endstate Orientation* by “concentrating on inherently telic verbs first and fixing end-state as head-of-event” (Hamann, 2015, p. 240) and are thus expected to master inherently telic and compositionally telic verbs with strong telicity markers from early on. In contrast, event expressions encoding further event types, such as states and processes, would emerge later in verb acquisition (Schulz et al., 2002). By contrast, “event-structural bootstrapping” is claimed to be unavailable for children with SLI (Penner et al., 2003), who are assumed to lack such an *Endstate Orientation*, resulting in qualitative differences in the acquisition of verbs, i.e., deviant development. According to Schulz (2018, p. 146), the weaknesses language-impaired children encounter with telicity arise “from an event-semantic representation of complex events in which the endstate as head-of-event is optional”. As a result, children with SLI are not capable of realizing that inherently telic (particle) verbs entail an endstate. Schulz and Kiese-Himmel (2006) investigated whether difficulties with telicity arise due to a global lexical deficit in children with SLI and found no meaningful differences in vocabulary size between children who mastered telicity and those who did not (see also Schulz, 2010). Moreover, no significant correlations emerged between performance on telicity tasks and vocabulary scores. The latter findings point to the existence of selective impairments in verb semantics¹¹¹, i.e., at the semantic level, that go beyond general lexical abilities (Schulz, 2018).

Sentential semantics

In addition to telicity, other aspects of semantics, such as universal quantifiers (Katsos et al., 2011) and the exhaustive interpretation of *wh*-questions (sentential semantics), have been shown to be impaired in children with SLI (Schulz & Roeper, 2011). SLI children’s difficulties with question interpretation are assumed to be related to their deficits in the production of *wh*-questions, which has been attested in several languages, e.g., English (van der Lely, 1994), German (Hamann et al., 1998), and French (Hamann, 2006, and

¹¹⁰ Schulz (2018) argues that the “initial preference for endstates is universal”; however, cross-linguistic variation exists with respect to the linguistic encoding of event components (p. 124).

¹¹¹ According to modular theories of language, semantics and lexicon constitute independent modules. As such, deficits in the domain of semantics can be considered as separate from lexical and pragmatic deficits (Grimm & Schulz 2021; see Chierchia & McConnell-Ginet, 1990 for definitions for semantics).

Prévost et al., 2011). Question interpretation requires knowledge of operator-variable dependencies and further presumes knowledge of which structures and contexts presuppose exhaustivity and, for multiple questions, knowledge of relations (lists of pairs), which are assumed to be compromised in children with SLI.

Research on the exhaustive interpretation of *wh*-questions in German revealed that exhaustivity in single and paired *wh*-questions is recognized around age five in monolingual typically developing children, whereas triple *wh*-questions are mastered by the age of six (Schulz & Roeper, 2011; see also Schulz, 2015). Comparisons with data from 415 five- and six-year-old TD children acquiring 19 different languages confirmed the cross-linguistic robustness of the acquisitional path of this structure in monolingual TD children: single *wh*-questions are mastered around age 5, while multiple *wh*-questions are mastered by age 6 (see Schulz 2010, 2013a, 2013b for cross-linguistic comparisons). This cross-linguistic uniformity was also reflected in the types of non-target responses, such as singleton answers to both single and multiple *wh*-questions and exhaustive subject or object lists as a response to multiple *wh*-questions. Such error patterns point to the children's initial inability to consider "underlying universal quantifier(s) (and mapping procedure)" when interpreting *wh*-questions (Schulz, 2015, p. 81).

Unlike their TD peers, children with SLI were found to have persistent problems, particularly with paired *wh*-questions up to the age of seven. In Wojtecka and Schulz's (2011) study, only 73% of the 5-year-old monolingual children with SLI showed mastery of exhaustivity in single *wh*-questions and performed below chance on paired *wh*-questions. Crucially, similar findings were reported for children with SLI of different language backgrounds, e.g., English (Roeper, 2004); French (Chondrogianni & Roesch, 2012); Serbian (Vuksanovic et al., 2012). These findings provide further evidence that children with SLI have deficits in the domain of semantics and not only with learning vocabulary, i.e., lexicon.

Of great importance is the finding that bilingual children with SLI are affected to a similar extent (Wojtecka & Schulz, 2011), though bilingual typically developing children have been reported to take a year longer than their monolingual peers (Schulz, 2013a) to achieve mastery of the interpretation of multiple *wh*-questions. BiSLI children were also reported to show persistent problems, especially with the exhaustive interpretation of multiple *wh*-questions (Grimm & Schulz, 2021; Lein et al., 2016; Schulz, 2013a). That bilinguals and monolinguals will be equally affected sounds plausible since once such knowledge is acquired, it should be manifest in both languages, at least if the

exhaustivity presupposition holds in both. Given its robust crosslinguistic path of acquisition and vulnerability in SLI, the exhaustivity of *wh*-questions appears to be a good candidate for teasing apart TD and SLI in bilingual contexts.

2.2.1.4 Executive functioning in children with SLI

Despite having normal-range¹¹² yet below age-average non-verbal IQ (Gallinat & Spaulding, 2014), children with SLI exhibit a range of weaknesses in executive functioning (EF). Numerous studies have shown that children with SLI underperform their typically developing peers on EF measures, including working memory, sustained attention, inhibitory control, and attention shifting (see Kapa & Plante, 2015; Montgomery et al., 2018b for overviews and Pauls & Archibald, 2016 for a meta-analysis). On the other hand, bilingual children are assumed to have an EF advantage over monolinguals, which has been contested by several researchers (for an overview, see Valian, 2015); see 2.1.3 for an in-depth discussion and definitions of the individual WM and EF constructs.

Working memory in SLI has been widely investigated within Baddeley's multi-component model (Allen & Baddeley, 2009; Baddeley, 2012; Baddeley et al., 2009, 2021), with a primary focus on storage capacities given their direct relevance to language development. Specifically, it has been argued that children with SLI "have a primary deficit in storing phonological representations, which interferes with language development and use" (Montgomery et al., 2018b, p. 24). In particular, phonological verbal short-term memory, i.e., the capacity of the phonological loop (measured by reduced item recall ability), has been reported to be vulnerable at different developmental stages and serves as a clinical marker in monolingual language-impaired children (Archibald & Gathercole, 2006; Bishop et al., 2006; Conti-Ramsden, 2003; Edwards & Lahey, 1998; Montgomery & Evans, 2009; Vugs et al., 2014). In a meta-analysis study, Graf Estes et al. (2007) found children with SLI to perform 1.27 standard deviations below their typically developing peers on measures assessing phonological/verbal short-term memory.

In addition to deficits in PSTM/VSTM, significant deficits have been reported for the central executive component of WM. Children with SLI frequently demonstrate

¹¹² Normal-range IQ means that non-verbal IQ is above the cutoff for cognitive impairment, i.e., $IQ \geq 85$ for SLI, and ≥ 70 for DLD.

reduced verbal working memory (i.e., complex verbal storage) abilities compared to their TD peers (Archibald & Gathercole, 2006, 2007; Briscoe & Rankin, 2007; Marton et al., 2014; Montgomery & Evans, 2009), although the precise memory-related mechanisms that constrain the capacity of verbal working memory are subject to debate. Some studies assume limitations in verbal WM to reflect dual deficits in phonological storage and aspects of impaired processing, e.g., poor controlled attention, particularly weak inhibitory control (Ellis Weismer et al., 1999; Marton et al., 2014), while others hold only storage deficits accountable for reduced VWM capacity. More recently, Archibald and Harder Griebeling (2016) have demonstrated that deficits related to phonological storage alone can account for deficits observed in complex storage-based verbal WM procedures in school-age learners with low language (LL) ability. Archibald and Harder Griebeling (2016) argued that traditional complex WM tasks, such as the backward digit recall, manipulate storage load but not processing demands, which remain constant throughout the task. By employing a novel complex WM span task that keeps storage demands constant (adjusted to the children's individual VSTM capacity) while varying processing load, no significant differences emerged between the LL group and their TD controls. In contrast, significant differences emerged between the groups when storage demands were manipulated while keeping the processing load constant, suggesting that "working memory impairments in SLI might be secondary to the well-documented phonological deficits in this group" (p. 252); see also Briscoe and Rankin (2007).

There is conflicting evidence concerning non-verbal memory deficits in children with SLI, with some studies reporting similar performance of SLI and TD children (e.g., Archibald & Gathercole, 2006) and others showing poorer performance in children with SLI relative to same-age peers (e.g., Bavin et al., 2005; Marton, 2008; Nickisch & von Kries, 2009). Vugs et al.'s (2013) meta-analysis study on visuospatial working memory in SLI revealed deficits with 0.63 SDs below TD controls in this area with medium to medium-large effect sizes (Cohen, 1992), which is indicative of a "clinically significant difference" between typically developing children and children with SLI (Pauls & Archibald, 2016, p.1082). However, the same meta-study revealed that the extent of verbal working memory deficits was twice to three times as high as that for visuospatial WM.

As to executive functions, existing evidence suggests that compared to their same-age typically developing peers, children with SLI demonstrate deficits across a range of domain-general EFs. In particular, language-impaired children are reported to perform

poorly on tasks tapping into inhibitory control (Marton et al., 2014; Spaulding, 2010; Vugs et al., 2014), auditory and visual sustained attention (Blom & Boerma, 2016; Boerma et al., 2017; Ebert & Kohnert, 2011; Finneran et al., 2009; Spaulding et al., 2008; Victorino & Schwartz, 2015), shifting, updating, and cognitive flexibility (Evans et al., 2018; Farrant et al., 2012). A meta-analysis study by Pauls and Archibald (2016) featuring 46 studies on cognitive flexibility and inhibitory control (overall age range: 4;0-14;0 years) found reliable differences in the latter two EF measures between TD children and children with SLI, which “remain present throughout development despite task demands or severity of linguistic impairment” (p. 1074). Similarly, a meta-analysis by Ebert and Kohnert (2011) revealed that children with SLI performed 0.69 SDs below TD controls on measures of sustained attention, a construct that resembles aspects of inhibition. As potential sources for poor inhibitory control, Pauls and Archibald (2016, p. 1082) mention “difficulty suppressing irrelevant information”, deficits in phonological short-term memory, and language deficits. Support for the role of language comes from Fatzer and Roeber’s (2012) study on 6 to 9-year-old children, which found that “[t]he extent to which articulatory suppression specifically impaired [executive function] performance coincided with the memory demand of the tasks” (p. 454). Interestingly, the impairment was more pronounced in 9-year-olds compared to 6-year-olds. In light of their findings, the authors argue that “the use of speech”, i.e., language, makes a substantial contribution towards the development of EF. Specifically, language is argued to play a crucial role in sustaining WM contents. Accordingly, the poor performance of children with SLI on inhibition tasks is likely to be related to their deficits in retaining verbal material, leading to weaker linguistic representations of the stimuli. Weak representations, in turn, are more susceptible to decay or interference, resulting in poor performance of children with SLI on inhibition measures relative to their unimpaired peers. Typically developing children, on the other hand, are capable of supporting memory-sustaining processes through reliance on a good quality “linguistic representation[s] of the task’s requirements” (Pauls & Archibald, 2016, p. 1082).

There has also been an increased interest in long-term memory in SLI, and emerging findings suggest that non-declarative implicit memory, i.e., procedural memory and learning are also vulnerable in SLI. Weaknesses in the procedural system are argued to manifest in SLI children’s difficulties in extracting “statistical dependencies from the input compromising lexical and grammatical acquisition” (Montgomery et al., 2018b, p. 27; see Evans et al., 2009; Lum et al., 2014). The latter deficits are captured in Ullman

and Pierpont's (2005) Procedural Deficit Hypothesis (PDH), which stipulates that language deficits in SLI stem "from an aberration in the neural network underlying procedural memory" (Kapa & Plante, 2015, p. 245) leading to grammatical deficits, which represent the hallmark of SLI. The PDH further posits that procedural learning deficits can lead to problems with other non-procedural components (such as WM), which depend "at least partly on the affected brain structures" (Montgomery et al., 2018b, p. 27). In contrast, declarative memory is assumed to be spared and may hence play a compensatory role not only in SLI but also in other neurodevelopmental disorders (Ullman, 2013; Ullman & Pullman, 2015). However, a large-scale study by Jackson et al. (2020) featuring one hundred 5-8-year-old children with SLI demonstrated that only visual declarative memory was unimpaired in SLI, whereas verbal declarative memory was impaired. Interestingly, procedural and verbal declarative memory deficits were no longer significant after WM and nonverbal IQ were statistically controlled for, suggesting that deficits in procedural memory "could largely be accounted for by working memory skills" (p. 4162).

Although numerous studies have documented an association between language development and executive function, the directionality and nature of the relationship are highly debated, i.e., it is still unclear whether EF deficits cause, contribute to, or co-occur with language deficits or whether disordered language acquisition impedes the development of executive function (Kapa & Plante, 2015; Marshall, 2020). Several studies within the limited input processing capacity framework investigated the role of WM as a potential causal factor for language difficulties attested in children with SLI (see Henry & Botting, 2017, for a review). Deficits in the verbal storage component of WM (VSTM/PSTM, the phonological loop) have been argued to be responsible for language deficits in SLI. For example, the phonological storage deficit hypothesis maintains that a "specific deficit in the temporary storage of novel phonological information underlies SLI" (Vugs et al., 2017, p. 3285; see Archibald & Gathercole, 2007; Bishop, 1996; Coady & Evans, 2008; for an overview see Henry & Botting, 2017). The latter storage component is implicated in the acquisition of the lexicon in typically developing children, i.e., vocabulary, particularly in the early stages of vocabulary development (Archibald, 2017; Gathercole & Baddeley, 1989; Leclercq & Majerus, 2010; Montgomery et al., 2010), and has been shown to be linked to vocabulary development in children with SLI as well (Gathercole & Baddeley, 1990; Montgomery, 2002). However, some studies showed that deficits in VSTM do not explain certain morphosyntactic problems. For

example, Riches (2012) found that the performance of 6-year-old children with SLI on the regular past tense inflectional suffix {-ed} in English and the complementizer *that* (Riches, 2012) was better accounted for by a syntactic priming task tapping LTM than by the Backward Color Span test, which taps VWM.

Moreover, many studies found associations between the verbal central executive component of WM and the processing of complex (morpho)syntactic constructions in children with SLI (Archibald, 2017; Delage & Frauenfelder, 2020; Frizelle & Fletcher, 2015; Montgomery & Evans, 2009; Noonan et al., 2014; Zebib et al., 2020b), which are reported to be most pronounced when computational demands are high (Archibald, 2017; Jakubowicz, 2011). For example, Montgomery and Evans (2009) found verbal WM (indexed by a complex-span task) to be positively correlated with comprehension for complex constructions that require computing a non-local dependency in children with SLI (age range: 6-12). More recently, Montgomery et al. (2016) proposed a WM-based account, within which WM limitations are directly linked to deficits in sentence comprehension in SLI (see also Montgomery et al., 2018a for an overview). It should be noted that the central executive component of WM has also been found to be associated with the development of language domains other than complex syntax, such as expressive and receptive vocabulary and verbal comprehension (Vugs et al., 2016, 2017).

An association with language deficits in SLI has also been reported for other EF functions, particularly for sustained attention (Blom & Boerma, 2016; Boerma et al., 2017). For example, Montgomery (2008) and Montgomery et al. (2009) found that auditory sustained attention explained 45% of the variance of the performance of children with SLI on online sentence processing. Sustained attention has also been shown to be associated with expressive vocabulary (Jongman et al., 2016) and narrative abilities (Blom & Boerma, 2016). Boerma et al. (2017) found the effects of language impairment on vocabulary and grammatical morphology performance to be mediated by auditory sustained attention in both monolingual and bilingual children aged between 5 and 8 years. The latter finding suggests that “the weakened ability of children with [S]LI to sustain their attention to auditory stimuli interfere[s] with how well incoming language is processed” (Boerma et al., 2017, p. 1).

To summarize, a reciprocal relationship seems to exist between SLI children’s linguistic deficits and the limitations in their “perceptual mechanisms that are important for language acquisition”, which “may lead to incomplete or inadequate processing of [incoming language] input” (Boerma et al., 2017, p. 3). As a consequence, substantially

more encounters with a particular structure or lemma (i.e., more exposure) would be necessary for it to be adequately processed and integrated into the language system of the child, resulting in a “protracted period of language development” and persistent language problems (Leonard, 2014a, p. 287).

As to bilingual children, a number of studies investigated whether working memory and executive function are impaired in bilingual children with SLI and how this interacts with their language development. Similar to their monolingual language impaired peers, simultaneous and sequential bilingual children with SLI have also been found to underperform their monolingual and bilingual typically developing peers on measures of VSTM, e.g., forward digit span, nonword repetition, and VWM, e.g., backward digit span (Armon-Lotem & Meir, 2016; Blom & Boerma, 2017; Girbau & Schwartz, 2008; Laloi, 2015; Thordardottir & Brandeker, 2013; Ziethe et al., 2013), as well as other EF measures such as auditory and visual sustained attention (Boerma et al., 2017; Ebert et al., 2014).

Concerning possible positive effects of bilingualism on WM capacities, most studies did not document an advantage for BiSLI children over MoSLI children, where both groups appear to perform equally poorly (Boerma et al., 2017; Engel de Abreu, 2011; Meir, 2017; Talli & Stavrakaki, 2020; Thordardottir & Brandeker, 2013), and some studies even reported a disadvantage (Laloi, 2015)¹¹³. For instance, Blom and Boerma (2017) documented positive effects of bilingualism on VSTM and VWM after controlling for vocabulary and morphology, but only when the bilingual typically developing and language-impaired groups were collapsed together as one group and compared to monolinguals, collapsing MoTDs and MoSLIs as well. Hence, whether the positive effect they observed applies to the BiSLI children cannot be ascertained.

As to the relationship between executive function and language development in BiSLI children, several studies reported associations between EF and language development, but the directionality of the relationship is still debatable. For example, Boerma et al. (2017) found that auditory and visual sustained attention mediated the impact of language impairment on vocabulary and morphology in bilingual children. As to WM, Meir (2017) compared the performance of preschool monolingual and Hebrew-Russian bilingual children with and without SLI on sentence repetition and two VSTM

¹¹³ Despite the lower performance of BiSLIs on VSTM in Laloi’s (2015) study, the interaction between bilingualism and SLI on performance on the forward digit span task was not significant implying that bilingualism does not aggravate WM deficits.

tasks with lower linguistic load, namely forward digit span and nonword repetition. Meir found that the significant difference between TD and SLI children on sentence repetition (which taps morphosyntactic abilities) remained visible even after controlling for VSTM. Hence, the author concludes that syntactic deficits in children with SLI are not solely caused by VSTM limitations in either monolingual or bilingual children. Moreover, Meir showed that the negative effect of VSTM on performance disappeared after controlling for low vocabulary scores. On the other hand, several studies have shown that WM limitations significantly contribute to poor performance of BiSLI children on measures assessing complex syntax in comprehension and production (e.g., Talli & Stavrakaki, 2020; Zebib et al., 2020b), though the exact nature of the relationship remains unclear.

2.2.2 Theoretical accounts for SLI

Theoretical accounts for SLI fall into two major classes regarding the assumed underlying cause for SLI: domain-general and domain-specific approaches¹¹⁴ (see Botting & Marshall, 2017, for a discussion). Accounts of the first type, such as the Generalized Slowing Hypothesis (Kail, 1994), capitalize on broad deficits in general or specific cognitive mechanisms, including working memory, auditory processing, and processing speed or capacity, as the underlying cause for SLI (e.g., Gathercole & Baddeley, 1990; Leonard et al., 2007; Montgomery et al., 2018b; see Henry & Botting, 2017; Kohnert et al., 2009; Leonard, 1998, 2014a, 2015; Montgomery et al., 2010; Reilly et al., 2014b; van der Lely, 2005b for discussions). They have been proposed in light of evidence showing that children with SLI demonstrate limitations in measures of verbal and non-verbal executive function alongside language deficits, especially deficits in verbal short-term memory as assessed by nonword repetition (Gathercole & Baddeley, 1990). Domain-specific approaches (linguistic representational or derivational accounts), on the other hand, posit that the underlying cause of SLI is a selective deficit in innate cognitive mechanisms specific to language processing and “concerns [aspects of grammatical knowledge, i.e.,] grammatical representations and derivations/[computational operations]”, independently of other general cognitive functions (Hamann et al., 2017, p.

¹¹⁴ These domain-general and domain-specific approaches to SLI are embedded in a larger debate on how the human brain is organized, i.e., whether a specialized cognitive module exists for processing linguistic information and how genetically endowed specialized cognitive systems devoted to language develop (Marshall et al., 2002; see Fodor, 1983; Levy & Kavé, 1999; Pinker, 1991, 1999).

142; see also Deevy & Leonard, 2004; Paradis et al., 2006; van der Lely, 2005a, 2005b). Hypotheses consistent with the domain-specific perspective (see 2.1.2.2 and 2.2.1.1 for elaborate descriptions of the hypothesis), such as the Agreement Deficit Hypothesis (Clahsen et al., 1997), the Extended Optional Infinitive Account (Rice & Wexler, 1996), and the Computational Grammatical Complexity hypothesis (van der Lely, 2004, 2005a) emphasize deficits in linguistic knowledge, or “problems with the computations underlying hierarchical, structurally-complex forms in one or more components of grammar” (Leonard, 2015; van der Lely, 2004, 2005b). Support for a domain-specific explanation for SLI comes from the widely documented deficits in tense and agreement morphology (Clahsen et al., 1997; Hamann et al., 1998; Rice & Wexler, 1996), phonology (Aguilar-Mediavilla et al., 2007; Bortolini & Leonard, 2000), and complex syntactic constructions, e.g., difficulties with accusative clitics in Romance languages, *wh*-questions, and relative clauses (Arosio et al., 2012; Delage & Frauenfelder, 2020; Friedmann & Novogrodsky, 2011; Hamann et al., 2017; Jakubowicz & Tuller, 2008; Tuller et al., 2011; van der Lely, 1998). The latter areas of grammar represent strong candidates for domain-specific systems because they involve aspects of morphosyntactic computation, such as “recursion [and] ... hierarchical non-local dependencies, not found in other cognitive domains” (van der Lely, 2005b, p. 53). Other language deficits falling within the scope of the domain-specific perspective of SLI include certain aspects of lexical acquisition (Froud & van der Lely, 2008) and interface phenomena between syntax and semantics, and aspects of pragmatics that depend on structural language knowledge (Andrés-Roqueta & Katsos, 2020; Bishop, 2000; Davies et al., 2016), such as referentiality (Fichman et al., 2022; van der Lely, 1997a), universal quantification (Katsos et al., 2011), telicity and exhaustive interpretation of *wh*-questions (Roeper, 2004; Schulz & Roeper, 2011). It is important to note that domain-specific accounts also assume interactions between WM limitations and aspects of computational complexity (Arosio et al., 2011; Jakubowicz, 2005, 2011), yet not a causal relationship.

A crucial difference between domain-specific accounts and accounts framed within the domain-general perspective (e.g., constructivist, usage-based, and functional approaches) that concerns language development in bilingual children is that domain-general accounts “assume the same general learning mechanisms for all learner populations” (Hamann et al., 2017, p. 142; see Ambridge & Lieven, 2011, for a critical review) and that knowledge of language is constructed by generalizing across the language input (Ambridge et al., 2008; Tomasello, 2003). Crucially, such accounts might

predict parallels between language development in typically developing bilingual children and monolingual children with SLI. From a limited processing capacity point of view, “incomplete processing [in SLI] is the functional equivalent of reductions in input frequency” experienced by bilingual children (Leonard, 2014, p. 289). Hence similarities are to be expected between monolingual children with SLI, who have EF weaknesses that would interfere with their language processing and uptake of language input, and BiTD children, whose language development is influenced by the divided nature of their linguistic input, which is often unbalanced and subject to changes over time (Grüter & Paradis, 2014; Hoff et al., 2012; Paradis & Jia, 2017). The temporary overlap between BiTD and MoSLI children (e.g., Håkansson & Nettelbladt, 1993; Paradis & Crago, 2004; Paradis et al., 2008; Rothweiler et al., 2012) prompted researchers to investigate whether bilingualism and SLI are “two of a kind” (Crago & Paradis, 2003; Paradis & Crago, 2004; see Paradis, 2010a and commentaries thereof, for in-depth discussions). However, qualitative analyses have demonstrated that the overlap between BiTD and MoSLI error patterns is rather superficial and that the quality and quantity of errors differ between the groups (Abed Ibrahim et al., 2018; Armon-Lotem, 2014; Armon-Lotem et al., 2008; Meir et al., 2016; Paradis & Crago, 2004; Roeper et al., 2001). Importantly, the temporary difficulties of BiTD children appear to manifest only in the early stages of bilingual acquisition and resolve with more exposure to the ambient language (Hamann & Belletti, 2006). Moreover, language acquisition research identified domains that are vulnerable in MoSLI children but not in BiTD children, namely complex syntax and phonological memory (Delage & Frauenfelder, 2020; Hamann et al., 2017; Thordardottir & Brandeker, 2013), making them promising areas for disentangling bilingualism from SLI.

Another important difference between domain-specific and domain-general accounts is that the latter predicts a “double delay” for bilingual children with SLI, which would particularly concern morphosyntactic properties. This is because BiSLI children are assumed to have weaker EFs like their monolingual language-impaired peers and, at the same time, be exposed to a reduced amount of input in each of their languages (see Paradis, 2010a, 2010b, for an elaborate discussion). Reduced input quantity is assumed to constitute an aggravating factor for BiSLI children, whose language uptake is compromised in the first place due to their WM and processing limitations. As a result, BiSLI children are expected to lag not only behind their BiTD peers but also behind their monolingual peers with SLI. This is captured by the Cumulative Effects Hypothesis (CEH, see Orgassa, 2009; Orgassa & Weerman, 2008; Paradis et al., 2017a), which claims

that BiSLI children are likely to never attain language abilities comparable to either MoSLI or BiTD children “for domains of morphosyntax known to be clinical markers in SLI and/or difficult to master for L2 speakers” (Paradis et al., 2017a, p. 882). In other words, the CEH hypothesis questions the dual language capacity of bilingual children with SLI. On the other hand, it has been argued that advantages brought upon by the bilingual experience, such as enhanced EF, “interdependence between bilinguals’ language systems, and more advanced learning mechanisms with older age of acquisition”, could alleviate “input/uptake” problems encountered by BiSLI children (Paradis et al., 2017a, p. 882; see Blom & Paradis, 2015; Ebert et al., 2014; Paradis, 2010a, 2010b; Peets & Bialystok, 2010). Although results of some cross-sectional studies lend support to the CEH (e.g., Orgassa & Weerman, 2008; Steenge, 2006; Verhoeven et al., 2011a, 2011b), numerous studies found no additive effects for bilingualism and language impairment on the acquisition of (morpho)syntax, where BiSLI children are reported to perform on par with their MoSLI peers (Blom and Boerma, 2017; Blom et al., 2013; Chilla et al., 2021; Gutierrez-Clellen et al., 2008; Morgan et al., 2013; Paradis et al., 2003, 2006, 2017a; Rothweiler et al., 2012; Tuller et al., 2018; see also studies reviewed in Kay Raining-Bird et al., 2016). Peña et al. (2020) found BiSLI children to be even more advantaged relative to MoSLI children on measures assessing morphosyntax and semantics. Concerning vocabulary, some studies report BiSLI children to be disadvantaged relative to their MoSLI peers (Blom & Boerma, 2017; Verhoeven et al., 2011b), while others do not (Degani et al., 2019). Although some might take this as evidence for a cumulative effect of bilingualism and SLI, it is important to consider the differences observed between MoSLI and BiSLI children in relation to the performance gap between MoTD and BiTD children (Degani et al., 2019). Since BiTD children frequently show poor performance on vocabulary measures (when assessed only in one of the child’s languages) relative to their MoTD peers (Kambanaros et al., 2014, 2015; Peña & Bedore, 2008), a similar pattern is expected for BiSLI relative to MoSLI. With respect to language development over time, some studies have shown that, unlike their typically developing bilingual peers who show enhanced language abilities with more exposure to the ambient language over time, additional input to the L2 is not correlated with better performance on morphological measures in BiSLI children (Blom & Baayen, 2013; Blom & Paradis, 2013, 2015). Such findings imply that the “resources for learning morphology (i.e., sensitivity to input properties, amount of input, L1 transfer, and age)

appear to be utilized [less, or perhaps] differently” by bilinguals with SLI than by bilinguals with typical language development (Paradis, 2016, p. 177).

In light of evidence showing enhanced cognitive control in bilinguals, Peets and Bialystok (2011) problematize dichotomizing “language ability as being controlled by either domain-specific or domain-general factors” (p. 315) and argue for a more inclusive model that encompasses both “representation and control components of linguistic knowledge, experience, and cognitive ability” (p. 318); see Botting and Marshall (2017) for a discussion on a Parallel Combination Approach. The authors stress that, in order to capture the complexity of SLI and bilingualism, more studies are needed that consider domain-specific and domain-general processes as separate contributors. Even though some of the data provided by the individual studies conducted in this thesis may be relevant to the debate, the dissertation adopts an applied focus and does not discuss which framework best accounts for bilingual SLI.

2.2.3 Long-term outcomes of SLI

Despite gains in language ability over time, a significant proportion of children with SLI still show persistent language deficits in later childhood, adolescence, and early adulthood (Beitchman et al., 1996; Botting, 2020; Brizzolara et al., 2011; Conti-Ramsden et al., 2001b; Johnson et al., 1999; McKinley & Larson, 1989; Tomblin, 2015). With increasing age, weaknesses with simple structures are usually overcome; however, vocabulary deficits (Beitchman et al., 2008) and difficulties with complex constructions and figurative language continue to persist (Leonard, 2014a; Nippold et al., 2009; Nippold & Fey, 1983). It appears that long-term outcomes are determined by the initial degree of severity and by the presence of comprehension deficits along with expressive difficulties in childhood (Conti-Ramsden et al., 2012; Law et al., 2008, 2009; Paul et al., 1983; Whitehouse et al., 2009). For example, Clegg et al. (2005) showed that individuals with severe receptive language impairment in childhood continued to exhibit impaired language functioning in their mid-30s. Even those individuals who appear to have good communication abilities and seem to have “grown out of it” can still encounter difficulties related to language and information processing in adolescence and adulthood (Durkin et al., 2011; Miller et al., 2008; Stothard et al., 1998).

Moreover, SLI brings about subclinical adverse consequences for other dimensions of development and can have “long-term ramifications in individuals’ lives

that go beyond language understanding and use” (Conti-Ramsden et al., 2018, p.238f; see Dubois et al., 2020 for a systematic review). In terms of nonverbal functioning, several long-term studies identified a decline in performance on nonverbal tasks in children with SLI from childhood to early adulthood, especially in individuals with persistent language difficulties (Conti-Ramsden et al., 2012; Krassowski & Plante, 1997; Paul & Cohen, 1984). In a study on 35 young adults with a history of language impairment (mean age: 21 years), Tomblin et al. (1992) reported an average decrease of approximately 10 standard scores in non-verbal IQ performance from childhood to early adulthood. Similar results were reported by Botting (2005), who followed 82 children diagnosed with SLI from age 7 to 14 and observed a significant drop of 23 IQ points over time (see also Johnson et al., 1999, and Mawhood et al., 2000 for similar findings). The latter finding could lead to diagnostic confounds, especially if assessment is conducted in later childhood, where poor language abilities might be mistakenly attributed to cognitive deficits rather than to an underlying language impairment.

A child diagnosis with SLI is further associated with a significant risk for developing reading impairments, lower literacy levels (Catts et al., 2014; Simkin & Conti-Ramsden, 2006), and poorer academic outcomes in adolescence and adulthood (Durkin et al., 2010; Snowling et al., 2001; Tomblin, 2008). A population-based cohort study by Law et al. (2009) found that, even after adjusting for distal sociodemographic risk factors, the likelihood for literacy problems in adulthood was almost twice as high (1.69) for children who met the criteria for SLI at kindergarten age than for children with typical language development. Furthermore, several studies have shown that SLI presents a significant barrier to academic attainment, employment, and financial independence (Tomblin, 2008; Winstanley et al., 2018a). In this respect, findings of long-term studies revealed higher school dropout and unemployment rates among young adults with a history of SLI compared to the general population (Botting et al., 2016a; Conti-Ramsden & Durkin, 2012; Elbro et al., 2011). Furthermore, individuals with a history of language impairment were found to be more prone to the risk of long-term unemployment (Law et al., 2009) and were more likely to hold less skilled occupations requiring lower educational levels, i.e., jobs categorized as lower in socioeconomic status (Blishen et al., 1987; Carroll & Dockrell, 2010; Conti-Ramsden et al., 2018; Johnson et al., 2010). In addition to the difficulties mentioned above, research has shown that a history of SLI increases the risk of mental and socio-emotional health problems (Botting et al., 2016a; Durkin & Conti-Ramsden, 2010), such as social adaptation difficulties, difficulties

establishing and maintaining peer relationships, depression, and social anxiety in adolescence (Botting et al., 2016b; Brownlie et al., 2016; Conti-Ramsden & Botting, 2008; Durkin & Conti-Ramsden, 2007; Durkin et al., 2017; Forrest et al., 2018; Snowling et al., 2006). For females, an elevated risk of experiencing sexual assault has also been identified, especially in cases where language difficulties were not identified in early childhood (Brownlie et al., 2017). Additionally, recent studies are reporting that youth with a history of SLI who were offered targeted intervention during early school years are less likely to have contact with the police and engage in aggressive rule-breaking behavior than their age-matched peers with unidentified language impairment (Bryan et al., 2015; Winstanley et al., 2018b). The latter findings highlight the “importance of timely assessment, early identification and subsequent targeted intervention for children with [SLI]” (Winstanley et al., 2018b, p. 257) for the crucial role this plays in ameliorating long-term adverse outcomes associated with SLI.

2.3 Diagnostic challenges in assessing bilingual children

Accurate assessment of language abilities is crucial to clinical decisions as to whether a bilingual child is given access to speech-language pathology services beyond general language support provided at educational facilities (Paradis et al., 2008). Such a “distinction is important since children with language impairment require different types of intervention from disadvantaged children whose inherent capacity for language is intact” (Roy et al., 2014, p.1). It also plays an integral role in informing therapy and tailoring intervention choices to meet the child’s needs (De Lamo White & Jin, 2011). In monolingual acquisition, poor performance on language measures and the presence of SLI clinical markers are assumed to reflect low language ability and thus indicate the presence of language impairment unless there are comorbidities or unusual limitations in the child’s language environment. However, bilingual typically developing children frequently demonstrate below-average language proficiency due to variations in language experience not only at a group but also at an individual level, e.g., in terms of linguistic background, timing, quality, and quantity of exposure to each of the languages, and language status (Chiat et al., 2013; Grüter & Paradis, 2014).

Moreover, clinical markers used to identify SLI in monolingual children might show temporary overlap with L2 interlanguage phenomena and are thus not always

suitable for reliable identification of SLI in bilingual contexts or “may have an unequal predictive value for different types of bilinguals” (Marinis et al., 2017, p. 266). For example, Rothweiler (2006) and Rothweiler et al. (2012) reported an overlap between MoSLI and BiTD children in SVA and the use of infinitives in German. In the same vein, Scherger (2016, 2018, 2019a) investigated the clinical utility of dative case marking as an indicator for SLI in simultaneous and early successive German-Italian bilingual children (age range: 4;0-7;0) in comparison to monolingual German children and found that BiTD children overused accusative case marking in dative contexts. In a subsequent study, Scherger (2022) compared the performance of eL2 children with and without SLI (mean age: 7;6 and 7;10, respectively) to a control group of monolingual TD children (mean age 7;3) and found that contrastive case marking did distinguish between BiTD and BiSLI in the latter age group, suggesting that case marking can be used for identifying SLI in bilingual children older than seven years of age. Although such error patterns are only temporary and are usually overcome over time in typical bilingual development (Paradis, 2010a, 2010b, 2010c; Scherger, 2018), using them as clinical markers to identify SLI in younger bilinguals would result in overdiagnosis with language impairment.

Due to the lack of assessment tools normed for bilingual populations (see Altman et al., 2022, for a discussion) and the partially overlapping linguistic profiles of bilingual children with typical language development and monolingual children with SLI, identification of true language impairment among bilinguals, especially sequential ones, is a very challenging endeavor (Arias & Friberg, 2017; Hamann, 2012). Since SLI is a disorder that concerns the process of language acquisition in general, it would affect both languages of a BiSLI child to the same extent (Kohnert et al., 2020; Paradis et al., 2021a), “with deficit manifestations modulated by the structure of each one” (M. Paradis, 2009, p. 176f). On the other hand, a bilingual typically developing child is likely to demonstrate an unbalanced command of her two languages, with superior linguistic abilities in at least one language, which often coincides with the child’s dominant language (Håkansson et al., 2003). Hence, to avoid cases of misdiagnosis, it has been recommended to evaluate a bilingual child at least in her dominant language (Fredman, 2006) and ideally in both of her languages (ASHA, 2023; IALP, 2011; RCSLT, 2007; see also Nayeb et al., 2021; Thordardottir, 2012, 2015a). However, this is rarely the case in clinical reality, and evidence-based recommendations for interpreting test scores of bilingual children are sparse (Peña et al., 2016). Besides, L1 assessment is often not feasible due to the lack of bilingual SLPs and adequate standardized language tests for (bilingual) children acquiring

their L1 in heritage contexts, especially for lower-incidence minority languages. Even if available, results may be unreliable due to differential attainment in L1 acquisition and/or L1-attribution¹¹⁵, which are often reported for heritage language speakers (Benmamoun et al., 2013; Montrul, 2008, 2016a). Not to mention that evaluation in two languages is time-consuming and that some of the immigrant L1 varieties undergo language change as a result of contact with the majority/societal language (L2), e.g., Immigrant Turkish in Germany (see Chilla, 2022; Chilla & Şan, 2017; Schröder & Dollnick, 2013). This leaves practitioners with no options but to resort to formative assessment in the L2. However, when assessing the child in only one language, it is crucial to consider the language experience of the child instead of measuring the language abilities of a bilingual child against the monolingual standard (Gathercole et al., 2013); otherwise, misdiagnosis is very likely to occur. Hence, as an alternative to assessment in both languages, it has been proposed to use (L2) tests with bilingual norm-referencing, if available. Yet, standardized tests normed with large bilingual data sets considering exposure patterns to all languages spoken by the child are very scarce (Bedore & Peña, 2008). The very few available exceptions were normed either based on samples of simultaneous bilinguals, such as the *Russian language proficiency test for multilingual children* (Gagarina et al., 2010), or early successive bilinguals like the German LiSe-DaZ (Schulz & Tracy, 2011), but not both. Hence, the norms, though bilingual, cannot be applied to all types of bilinguals. Moreover, tests designed for bilinguals of a particular L1-L2 combination might not be suitable for bilinguals acquiring the language of assessment alongside a different language, as “this might lead to different developmental profiles” (Armon-Lotem, 2018, p. 621). For instance, the *Russian language proficiency test for multilingual children* was normed with Russian-German bilinguals residing in Germany and might thus not be suitable for assessing Russian-Hebrew bilinguals given the typological differences between German and Hebrew. This might result in differences in cross-linguistic influence on Russian, which, in turn, can affect performance on the test and render the bilingual norms not applicable (Armon-Lotem, 2018).

Based on research on simultaneous preschool French-English bilingual children (Thordardottir, 2011; 2014; Thordardottir & Brandeker, 2013), Thordardottir observed that the children’s scores on monolingual tests assessing vocabulary and grammar varied

¹¹⁵ Language attrition is defined as “the non-pathological decrease in proficiency in a language that had been previously acquired by an individual” (Köpke & Schmid, 2004, p. 5), i.e., following complete acquisition, the linguistic system undergoes some sort of loss.

as a function of the amount of exposure they received to each of their languages, where even those with an exposure ratio of 75% to the language of assessment scored 0.25 to 0.5 standard deviations below the monolingual mean. Since -1 to -1.25 SD is viewed as the cutoff score separating typical from impaired language development in monolinguals (Leonard, 2014a), she proposed adjusting the monolingual cutoff criteria for performance in the pathology range according to the status of the language being tested as the dominant or non-dominant language when monolingually normed tests are used to assess bilinguals. This procedure would give bilingual children some slack by lowering the cutoff criteria to account for variation characteristic for bilingualism. The proposed differential cutoff criteria are summarized in Thordardottir (2015a, p. 343) and are as follows: children assessed in their weaker language are allowed to score between 2.25 and 2.5 SD below the monolingual mean before they are considered to be language-impaired, balanced bilinguals between -1.75 and -2 SD, and for L2-dominant bilinguals, the cutoff point lies between -1.5 SD and -1.75 SD. At the same time, Thordardottir calls for caution when applying the proposed cutoff criteria, as they are based on simultaneous bilinguals up to age five and might not apply to other types of bilinguals. For example, Hamann et al. (2020) showed that despite applying dominance-adjusted cutoff scores, almost the entire bilingual sample, consisting of early and late successive bilinguals, scored in the pathology range on L2 expressive and receptive vocabulary measures (WWT, Glück 2011). In sum, none of the proposed diagnostic solutions based on standardized assessments appears to provide a satisfactory solution to the assessment problem in bilingual contexts; see De Lamo White and Jin (2011) for a discussion of different biases associated with standardized tests, e.g., content bias and linguistic bias. This makes it crucial to develop reliable (L2) language screening tools that could disentangle the effects of bilingualism and language impairment in bilingual contexts, which are also easy and fast to administer.

2.3.1 LITMUS tools for bilingual language assessment¹¹⁶

In a collaborative effort to cope with the diagnostic challenges in multilingual contexts, a battery of tools was designed within the international research network COST Action

¹¹⁶ Parts of this section are based on Abed Ibrahim, Lina, & Fekete, István (2019). What machine learning can tell us about the role of language dominance in the diagnostic accuracy of German LITMUS non-word and sentence repetition tasks. *Frontiers in Psychology*, 9, 2757. <https://doi.org/10.3389/fpsyg.2018.02757>

IS0804, which features more than 30 countries ‘*Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment*’ (www.bi-sli.org). The tools were constructed according to a set of linguistic principles that allow cross-linguistic comparability. At the same time, they were devised to minimize the effect of factors related to bilingualism so that SLI can be reliably identified in bilingual children with different language combinations. These tools are known as the LITMUS tools (Language Impairment Testing in Multilingual Settings, see Armon-Lotem et al., 2015a; Armon-Lotem & Grohmann, 2021). The latter tools include tasks targeting specific language areas known to be vulnerable in children with SLI, such as contrastive case marking (LITMUS-CCM; Ruigendijk, 2015), subject-verb agreement (LITMUS-SVA; de Jong, 2015), and exhaustive interpretation of *wh*-questions (sentential semantics; Schulz, 2015), as well as tasks assessing broader language abilities such as sentence (SRTs) and nonword repetition tasks (NWRTs), which tap morphosyntactic and phonological abilities (depending on the construction of the nonwords), respectively (Marinis & Armon-Lotem, 2015; Chiat, 2015). The LITMUS toolkit further includes the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015), which enables practitioners to gather relevant background information that helps them quantify the bilingual experience of a child and estimate individual risk factors for language impairment.

The studies included in the current thesis focus on sentence repetition and nonword repetition since such tasks have been shown to reliably identify SLI in monolinguals (Conti-Ramsden et al., 2001a) and to be less reliant on prior language experience compared to other language measures in bilinguals, e.g., expressive/receptive vocabulary (Chiat et al., 2013; Thordardottir & Brandeker, 2013). Both tasks are assumed to mirror language production mechanisms in that nonword repetition requires access to underlying lexical and sub-lexical representations (Coady & Evans, 2008; Dollaghan & Campbell, 1998), whereas sentence repetition involves conceptual processing and grammatical encoding, including the assembly mechanisms necessary for combining lexical entities into grammatically sound and meaningful sentences (see Nag et al., 2018 for a discussion).

Depending on their construction, SRTs and NWRTs can be designed to not only assess (phonological) working memory (Archibald & Gathercole, 2006; Archibald & Joanisse, 2009) but also the command of syntactic and phonological representations/derivations (see Klem et al., 2015; Polišenská et al., 2015 for sentence-repetition; Gallon et al., 2007 for non-word repetition). Such linguistic

representations/derivations, particularly their complexity, have been found to crucially affect performance, especially in language-impaired populations (e.g., Ferré et al., 2012; Friedmann et al., 2015), so that it has been argued that they are not mere measures of working memory (Polišenská et al., 2015; Vinther, 2002). Thanks to their versatility, they are ideal for targeting language-specific and cross-linguistically challenging (morpho)syntactic and phonological structures while minimizing avoidance strategies (see Hamann et al., 2017 for SRT). Hence, they were selected for subsequent cross-linguistic development within and beyond COST Action IS0804 (Chiat, 2015; Marinis & Armon-Lotem, 2015). Multiple recent studies had reported good to excellent diagnostic accuracy for LITMUS sentence and nonword repetition tasks in simultaneous and early successive bilinguals with different L1 backgrounds, where the best diagnostic accuracy was obtained when the tasks were used in combination (Abed Ibrahim & Fekete; 2019; Abed Ibrahim & Hamann, 2017; Armon-Lotem & Meir, 2016; Chiat & Polišenská, 2016; Chilla et al., 2021; de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018; Zebib et al., 2020b). Abed Ibrahim et al. (2020), Hamann et al. (2020), and Hertel et al. (2022) have further shown that using the tasks combined can help avoid cases of overdiagnosis not only in heritage but also in refugee school-aged children who started acquiring the L2 after the age of five, i.e., late successive child bilinguals. This is a very encouraging outcome, especially since the tasks are easy and fast to administer and could serve as first screening tools in bilingual contexts.

2.3.1.1 Sentence repetition tasks

Sentence repetition, also known as immediate sentence verbatim recall or sentence imitation, measures a child's ability to repeat verbatim a series of sentences that she heard, i.e., the "preservation of the just-heard phonological form of the lexical items and their linear sequence" (Polišenská et al., 2015, p. 115). Compared to other types of language measures, it is less constrained by pragmatic and discourse factors (Hamann et al., 2017; Polišenská et al., 2015) and is widely used in clinical assessment as a measure of sentence-level abilities (e.g., for English: TOLD-P: 5, Newcomer & Hammill, 2019; CELF-5, Wiig et al., 2013, for German: CELF-5, Wiig et al., 2020; SETK 3-5, Grimm, 2001; SSV, Grimm, 2003). Moreover, unlike many expressive and receptive language tests, sentence repetition appears to be capable of detecting "residual language processing weaknesses"

(Klem et al., 2015, p. 146; Stothard et al., 1998) and can predict prospective persistent expressive language deficits in young children (Everitt et al., 2013).

Since sentence repetition also recruits working memory resources (Alloway & Gathercole, 2005), the task has frequently been employed to assess WM (e.g., Delcenserie et al., 2013; Meir, 2017; see Zebib et al., 2020b for a discussion). However, sentence repetition is not mere retrieval of “an episodic, form-based representation of the sentence” (Nag et al., 2018) as it requires processing the incoming input string, storage, analysis, and reconstruction thereof, especially when the sentences are long enough to prevent mere phonological reiteration (Allen & Baddeley, 2009; Marinis & Armon-Lotem, 2015; see also *The Regenerative Hypothesis*: Lombardi & Potter, 1992; Potter & Lombardi, 1990). Within the working memory framework, sentence repetition is constrained by the capacity of the episodic buffer¹¹⁷ rather than the phonological loop (Baddeley, 2000). The episodic buffer constitutes a distinct memory component that is “uniquely linked with language skills” (Alloway & Gathercole, 2005, p. 279) and is assumed to have a causal effect on the development of language abilities in children (Klem et al., 2015, p. 147).

From a language processing perspective, sentence repetition has been argued to draw upon “a broad range of language processing skills” (Klem et al., 2015, p. 148) and thus comprises a complex linguistic task that involves several aspects of linguistic knowledge, including morphosyntactic, lexical-semantic and phonological processes (Alloway & Gathercole, 2005; Moll et al., 2015; Riches et al., 2010). The significant reliance of sentence repetition on language skills has been established in various studies, especially when the stimuli's length exceeds the phonological loop's capacity. For example, Polišenská et al. (2015) investigated the influence of different types of LTM knowledge on sentence repetition performance in preschool monolingual English and Czech-speaking children and found performance to be highly dependent on long-term linguistic representations, particularly morphosyntax and lexical phonology with function words of particular significance. Given the well-documented difficulties in repeating sentences with more complex syntactic structures (e.g., relative clauses, *wh*-questions, and passives) relative to simpler ones (e.g., Riches, 2012), sentence repetition is argued to specifically tap (morpho)syntactic representations beyond other aspects of linguistic knowledge. An effect of syntactic complexity has been reported for monolingual

¹¹⁷ As outlined in 2.1.3, the episodic buffer was added to Baddeley's original multicomponent WM model in order to accommodate LTM's contribution (Allen & Baddeley, 2009; Baddeley et al., 2009).

language-impaired adults (Poll et al., 2010), adolescents (Riches et al., 2010), and, more importantly, typically developing and language-impaired children (Hamann et al., 2017; Riches, 2012; Szterman & Friedmann, 2015). In fact, poor performance on sentence repetition has been identified as the most reliable clinical marker for the identification of SLI in monolingual English-speaking children compared to other clinical markers like past tense marking (e.g., Alloway & Gathercole, 2005; Archibald & Joanisse, 2009; Conti-Ramsden et al., 2001a; Redmond et al., 2019; Riches, 2012; see also Marinis & Armon-Lotem, 2015). Sentence repetition tasks have also been shown to have good to excellent diagnostic accuracy for identifying SLI in children from a variety of languages other than English, e.g., Arabic: Shaaan (2010); Taha et al. (2021b); Hebrew: Armon-Lotem and Meir (2016); French: Leclercq et al. (2014); Thordardottir et al. (2011); see Rujas et al. (2021) for a scoping review. The latter promising results raise the question of whether sentence repetition can be reliably used for identifying language impairment in bilingual contexts.

Compared to other aspects of language, like morphology and lexicon, syntactic complexity appears to be more robust against the effects of bilingualism in typical bilinguals. Unlike the problems encountered by children with SLI, bilingual children with typical language development may exhibit difficulties with feature assembly (Lardiere, 2008) and uninterpretable features (Tsimpli & Dimitrakopoulou, 2007) or might even avoid complexity (Tuller et al., 2015). They are, however, assumed to have an intact language faculty and WM. Thus, having been acquired in the L1, syntactic operations such as recursion¹¹⁸ (see de Villiers et al., 2020 for a recent discussion on recursion), embedding, and (*wh*-)movement do not have to be acquired again and should not be problematic for them given sufficient exposure to the L2 (Roeper, 2011; Schwartz, 2004; see also Scheidnes, 2020). Hence, SRTs incorporating aspects of syntactic complexity represent good candidates for adaptation into an assessment tool for bilingual populations in the absence of other formative assessment tools, especially since they are fast and easy to administer.

Studies looking into the influence of bilingual experience-related factors on performance on SRTs showed that, compared to other language measures, such as lexical

¹¹⁸ Syntactic recursion exists “whenever a grammar can freely embed a phrase XP inside the same phrase XP” (Widmer et al., 2017, p. 801). Widmer et al. (2017) add that “an XP_n is recursively embedded into an XP_{n+1} if the distributional properties of XP_n and the relationship between XP_n and XP_{n+1} are the same for all n , and there is no grammatical restriction on the value of n .” (p. 801).

tasks, SRTs appear to be less affected by exposure variables, such as AoO and LoE (e.g., Thordardottir & Brandeker, 2013). Several studies found early successive bilinguals to perform on par with monolinguals within two years of exposure to the societal language (e.g., Armon-Lotem et al., 2011). However, when assessed in both languages, bilingual children tend to perform better in their dominant language (Meir, 2018; Verhoeven et al., 2012). Moreover, upon comparing different sentence repetition tasks, Chiat et al. (2013) found performance on SR to be adversely affected by the lexical complexity of the test items. Nevertheless, Chiat et al. (2013) point out that despite being susceptible to the impact of quantitative aspects of L2 input and vocabulary, bilingual children tend to catch up with their monolingual peers on SR in a relatively shorter time than on other types of linguistic tasks. Hence, they recommend using syntactically age-adequate sentences but controlling for lexical demands (using basic early-acquired vocabulary) and minimizing language-specific effects in lexical phonology (see also Armon-Lotem, 2018, and Polišenská et al., 2015). This was pursued during COST Action IS0804 and resulted in developing the so-called LITMUS sentence repetition tasks in multiple languages following the same construction principles (LITMUS-SRep; Marinis & Armon-Lotem, 2015; see also <https://www.litmus-srep.info/>). The tasks developed within the LITMUS-SRep framework incorporate syntactically complex structures known to be cross-linguistically vulnerable in language-impaired children, in addition to a set of structures that serve as clinical markers for SLI in a particular language.

The German LITMUS-SRT (Hamann et al., 2013) under investigation in this thesis was constructed according to the LITMUS-SRep principles (Marinis & Armon-Lotem, 2015) and builds on the notion of computational complexity. Within the generative framework, computational complexity can be determined by the number and nature (e.g., merge vs. movement, distance of dependencies, and depth of embedding) of syntactic operations necessary for deriving a syntactic structure (Friedmann et al., 2009; Gibson, 1998; Hamann et al., 2007; Jakubowicz, 2005; Jakubowicz & Tuller, 2008). Children with atypical language acquisition are proposed to have a greater deficit in constructions with a higher degree of computational complexity, as the latter are more taxing to working memory resources (Chomsky, 2005; Hamann et al., 2007; Jakubowicz & Tuller, 2008; see Delage & Frauenfelder, 2019 for a discussion). A particular difficulty for children with SLI has been reported for structures involving movement and intervening elements between the source of the moved constituent and its landing site, e.g., object which-questions and object relative clauses with a lexical subject (Friedmann

et al., 2015; Rizzi, 2004; see **2.1.2.2** for a deeper discussion). Accordingly, the German LITMUS-SRT incorporates a set of syntactically complex, i.e., computationally more demanding structures identified as difficult for children with SLI cross-linguistically, in addition to a set of structures reported to be challenging for German MoSLI children, such as topicalization and the sentence bracket, which represent crucial milestones in the (L2) acquisition of German word-order properties (see Clahsen, 1986b; Griebhaber, 2013; Haberzettl, 2005; Wegener, 1992). The complex structures involve computational operations like syntactic movement (measured, for example, by the number of overt movement operations), in particular *wh*-movement, i.e., fronting of interrogative or relative pronouns (Hamann et al., 1998; Jakubowicz, 2011; Marinis & van der Lely, 2007; van der Lely, 1998), and/or clausal embedding, e.g., relative clauses (Friedmann & Novogrodsky, 2011; Hamann & Tuller, 2014; Scheidnes & Tuller, 2019). It has been shown that LITMUS-SRTs eliciting structures involving the latter operations can be reliably used to tease apart typically developing bilingual children from monolingual and bilingual children with SLI, not only in bilingual but also in bilectal settings (e.g., Armon-Lotem & Meir, 2016; Meir et al., 2016, 2017 for Russian and Hebrew; de Almeida et al., 2017; Fleckstein et al., 2018; Zebib et al., 2020b for French; Abed Ibrahim & Hamann, 2017; Abed Ibrahim et al., 2018; Hamann & Abed Ibrahim, 2017; Hamann et al., 2017; for German; Theodorou et al. (2017) for Cypriot-Greek; see also Marinis et al., 2017 for an overview).

Importantly, Armon-Lotem and Meir (2016) showed that although the highest level of diagnostic accuracy can be achieved using a combination of SRTs in the child's L1-Russian and L2-Hebrew (applying bilingual cutoffs), good diagnostic accuracy can still be achieved if SRT is only administered in the societal language (L2-Hebrew). In the same vein, Abed Ibrahim et al. (2018) and Chilla et al. (2021) looked into the potential influence of L1-L2 typological differences on the performance of bilingual children with Arabic, Portuguese, and Turkish as L1s and German or French as L2 on German and French LITMUS-SRTs, respectively. L1 influence surfaced neither in the overall performance nor in the performance on the individual structures included in the task, nor in the expected L1-driven error patterns, confirming the applicability of the task to bilingual children with diverse L1 backgrounds. It should be noted, however, that most of the studies on LITMUS-SRT report lower-cutoff scores separating TD from SLI in bilingual children, including the studies mentioned above. Furthermore, the task appears to be unsuitable for assessing bilingual children with less than 12 months of exposure to

the L2 (see Tuller et al., 2018). More recently, Hamann et al. (2020) investigated whether the German LITMUS SRT can identify typical language development in recently arrived late successive bilingual Syrian refugee children who started acquiring German after age five. For this purpose, they compared the performance of 15 typically developing school-age Syrian refugees (age range: 6;6-12;8 yrs.; LoE: 18-41 months) to 12 typically developing simultaneous and early successive bilingual German-Arabic speaking children (age range: 6;0-12;9 yrs.; LoE: 46-120 months), i.e., heritage bilinguals, on the German LITMUS-SRT (Hamann et al., 2013; <https://osf.io/8kb4n/overview>) and QU-NWRT (Grimm et al., 2014). Their results showed that while both bilingual groups performed similarly on the LITMUS-QU-NWRT regardless of age and input variables, only those refugee children with more than 24 months of exposure to the L2 performed on a par with their heritage peers on LITMUS-SRT. In fact, not only did refugee children with fewer than 24 months of exposure to German perform significantly worse than heritage bilinguals, but they also performed below the cutoffs for SLI previously established for younger heritage bilinguals (cf. Hamann & Abed Ibrahim, 2017). In sum, German LITMUS-SRT offers a reliable tool for the identification of language impairment in 2L1 and eL2 children; however, in cases of limited L2 exposure or older age of onset to the L2, it is recommended to complement LITMUS-SRT with the LITMUS-QU-NWRT, given its relatively small linguistic load and robustness against the influence of L2 experience (Hamann et al., 2020).

2.3.1.2 Nonword repetition tasks

Non-word repetition (NWR) assesses the ability to repeat strings of sounds that may or may not resemble existing words in a given target language (Armon-Lotem, 2018, p. 629). NWR belongs to the core assessment measures used for diagnosing language impairment (Chiat, 2015) and has been identified as a reliable clinical marker of SLI in monolingual children from a variety of typologically different languages, e.g., Arabic (Khater, 2016; Saiegh-Haddad & Ghawi-Dakwar, 2017; Shaalan, 2020), English (Bishop et al., 1996; Coady & Evans, 2008; Conti-Ramsden et al., 2001a; Gathercole, 2006a, 2006b), French (Leclercq et al., 2013; Thordadottir & Brandeker, 2013), Italian (Casalini et al., 2007; Dispaldro et al., 2013), Russian (Kavitskaya et al., 2011) and Spanish (Gibson et al., 2015; Girbau, 2016; Girbau & Schwartz, 2007). An advantage of NWR over other language measures is that it is less affected by prior knowledge of vocabulary and morphosyntax

(Chiat, 2015; Thordardottir & Brandeker, 2013) and counts as a relatively culturally fair measure, which could be used for the assessment of children with diverse linguistic and socioeconomic backgrounds (Chiat, 2015; Chiat & Polišenská, 2016; Engel et al., 2008). As such, NWR tasks offer promising tools for the identification of SLI, especially in bilingual children with limited exposure to the L2; see Schwob et al. (2021) and Ortiz (2021) for meta-analyses on nonword repetition tasks in bilingual contexts.

Measured by increasing numbers of syllables, NWR has traditionally been used to assess phonological working memory (e.g., Archibald & Gathercole, 2006; Bishop et al., 1996; Gathercole & Baddeley, 1990; Montgomery, 2004). A classic finding of research in this vein is that children with SLI encounter greater difficulty with longer nonwords, i.e., three- to five-syllable sequences, relative to shorter ones. Such findings are usually interpreted in terms of limitations in phonological working memory capacity in children with SLI since accuracy is high on two-syllable nonwords but sharply declines as the nonword length increases (Gathercole & Baddeley, 1990). However, the ability to repeat nonwords is not only constrained by phonological working memory but also requires several phonological skills, which are assumed to relate to the capacity of learning new words in early childhood (see Gathercole, 2006a, and commentaries thereof for a review; Melby-Lervåg et al., 2012). Such skills include speech perception, phonological encoding, also known as phonological sensitivity¹¹⁹ (Metsala, 1999), storage and retrieval of phonological representations, phonological assembly (Gathercole & Baddeley, 1990; James et al., 1994), and articulation. Although each of these skills can be deficient in children with atypical language development (see Coady & Evans, 2008, for a review; Marshall, 2014), articulatory difficulties are reported to play a rather limited role in explaining poor performance in nonword repetition in children with SLI (see Archibald et al., 2013, for a discussion). Several studies have shown that monolingual children with SLI are not only sensitive to the amount of phonological material, i.e., the number of syllables in the non-words, but also to phonological complexity. For example, the presence of consonant clusters, especially in the form of branching onsets or a consonantal coda, has been reported to comprise a particular source of difficulty for children with phonological impairments in general and (phonological) SLI in particular (for English, see Barlow, 2001; Briscoe et al., 2001; Marshall et al., 2002; Marshall &

¹¹⁹ “Children with SLI might have a deficit in phonological encoding, in which inefficient or inaccurate encoding of phonological material might [...] result in a less robust or more quickly decaying phonological trace, thereby limiting repetition accuracy” (Coady & Evans, 2008, p. 15).

van der Lely, 2009; Tamburelli & Jones, 2013; for French, see Ferré et al., 2012; Gallon 2002; Gallon et al., 2007; for German, see Ott et al., 2006; Romonath, 1991). A number of studies even reported that language-impaired children encountered more problems with shorter complex items relative to longer simpler ones with a tendency to simplify complex syllables into simple CV syllables in monosyllabic words (e.g., Gallon, 2002; Marshall et al., 2002).

Designing a nonword repetition task that identifies language impairment in bilingual children without disadvantaging those with less exposure to the target language is not straightforward. Despite being less reliant on language-specific knowledge compared to other types of tasks, there is substantial evidence that performance on NWR (both within and across languages) is not only affected by item length and complexity of the non-words but also by other properties such as word-likeness¹²⁰, prosodic structure, phonotactic probability¹²¹, and neighborhood density¹²². For instance, children are found to perform significantly better on non-words that are more wordlike, carry language-specific stress patterns, contain language-specific morphemes, or have higher phonotactic probability (Casalini et al., 2007; Coady et al., 2010; Dollaghan et al., 1995; Edwards et al., 2004; Jones et al., 2010; Leclercq et al., 2013; Messer et al., 2010; for an overview, see Chiat, 2015). Munson et al. (2005) and Rispen et al. (2015) examined the influence of phonotactic probability on nonword repetition in monolingual English and Dutch-speaking children, respectively. They found that phonotactic probability effects on nonword repetition were more pronounced in younger typically developing children and significantly correlated with their vocabulary levels. In contrast, no such effects were observed for older typically developing school-age children. These findings suggest that the influence of phonotactic probability on nonword repetition is very likely to be mediated by vocabulary size, which is subject to developmental changes “perhaps due to

¹²⁰ Word-likeness refers to “the extent to which as sound sequence is typical of words in a language” (Bailey & Hahn, 2001, p. 568).

¹²¹ Phonotactic probability is defined as the relative positional frequency of phonemes (e.g., the frequency of /p/ occurring in the second position of the word), and of phonemes sequences within words, (e.g., bigrams as in the frequency of /s/+p/, or trigrams as in the frequency of /s/+p/+t/ (Chiat, 2015, p. 130; Leonard, 2014a, p. 63). It should be mentioned that “phonotactic probability does not take into account the position of phonemes within the prosodic and syllabic structure of the word”, i.e., phoneme sequences might be the result of syllable contact (Chiat, 2015, p.130).

¹²² Neighborhood density refers to the number of words differing from a given word by one segment by means of deletion, substitution or addition (Leonard, 2014a; Luce & Pisoni, 1998). Hence, a word like *sit* appears to have many neighbors (*sip*, *site*, *seat*, *hit*, *spit*, *it*, among others), while a word like *syntax* is unlikely to have any neighbors.

its influence on the ongoing refinement of phonological categories” (Munson et al., 2005). Taken together, this implies that non-word repetition ability is also affected by long-term memory (LTM) verbal knowledge, i.e., experience and knowledge of lexical phonology and semantics, which, depending on the nature of the non-words, is generally shown to relate to expressive and receptive vocabulary levels in monolinguals (Chiat & Polišenská, 2016; Edwards & Lahey, 1998; Gathercole, 2006a). Similar results were reported for bilingual children who seemed to perform on par with their monolingual peers once vocabulary was controlled for (Engel de Abreu et al., 2013). These findings become particularly relevant when distal factors known to modulate the child’s linguistic experience, such as the family’s socioeconomic status, are considered. Children from high SES households tend to be exposed to enriched input and more differentiated vocabulary compared to their peers with low SES (e.g., Fernald et al., 2013; Hoff, 2003a, 2003b). Hence, they are likely to have a larger vocabulary size and to be more familiar with morphemes from morphologically complex vocabulary, which are sometimes integrated into the stimuli of an NWR task as items with low phonotactic probability. As a result, “more lexical representations (including words with less phonotactic sequences) are available to facilitate the recall of a nonword” (Rispen et al., 2015, p. 88). Such socioeconomic status-based differences in vocabulary size might influence performance on nonword repetition tasks and should be taken into consideration in task construction since “[n]ot only will there be a confound with the language elements in the tests, but there could also be a confound of factors such as bilingualism with SES” (Hamann, 2017, p. 1315).

Different LITMUS nonword repetition tasks manipulating factors known to influence performance on nonword repetition, such as length, prosody, and/or syllable complexity, were constructed within the COST IS0804 framework for NWR (see Chiat, 2015; <https://www.bi-sli.org/nonword-repetition>). Of particular relevance for this thesis is the German LITMUS Quasi-Universal-Nonword-Repetition task (LITMUS-QU-NWR, Grimm et al., 2014), which was constructed in close parallel to the French LITMUS-QU-NWR (dos Santos & Ferré, 2018). The task has also been recently adapted into other languages such as Portuguese, Dutch, and Lebanese Arabic (see <https://www.bi-sli.org/qu-nonword-repetition>). Unlike the LITMUS crosslinguistic (CL) NWR task (e.g., the Dutch CL-NWR task Boerma et al., 2015; Boerma & Blom, 2017; <https://www.bi-sli.org/cl-nonword-repetition>), which primarily tests phonological short-term memory and consists of phonologically simple non-words compatible with the

phonological properties of any language, the German LITMUS-QU-NWRT was devised to tap more directly into phonological abilities by focusing on phonological complexity. LITMUS-NWRTs of this type systematically vary segmental (articulatory difficulty), syllabic (presence or absence of clusters), and sequential complexity (types of consonant and syllable sequences), combining them into non-words of increasing phonological complexity. At the same time, to avoid penalizing bilingual children, language-specific phonological properties are controlled as far as possible, e.g., by avoiding “the use of language-specific phonotactic patterns” (Grimm & Schulz, 2021; Messer et al., 2010). In order to limit the effects of lexical knowledge, the non-words of the German LITMUS-QU-NWRT were constructed to be maximally distinct from existing words in German and were created using elementary blocks (segments and syllables) that are cross-linguistically well-attested (Maddieson et al., 2011). In line with the COST Action IS0804 framework (Chiat, 2015), the latter blocks were combined and manipulated in two sets: a set of phonologically complex items with phonological properties common in most of the world’s languages (the language-independent part, LI), and another set of items using the same building blocks of the LI part in addition to the combination of /f, s/ and an obstruent in word-initial and final positions as a complexity variable specific to German and some other languages (the language dependent¹²³ part, LD). The maximum non-word length is limited to three syllables in both parts to minimize the working memory load, which could undermine the effect of phonological complexity. Various studies reported adverse effects of language-specific properties of the NWRTs on the performance of bilingual children resulting in insufficient diagnostic accuracy, e.g., Armon-Lotem and Meir (2016), Boerma et al. (2015), Kohnert et al. (2006), and Windsor et al. (2010). However, since the construction of the LD part of the German LITMUS-QU-NWRT considerably differs from other language-specific NWRTs (see 3.2.3.2), bilingual children are not expected to be disadvantaged by the LD part of this particular task. Although structures with higher phonological complexity are generally more error-prone in TD children, they appear to be “disproportionately difficult” for children with SLI (Chiat, 2015, p. 137) who struggle with phonological complexity (Archibald & Gathercole, 2006; dos Santos & Ferré, 2018; Jones et al., 2010). Thus, assuming that monolingual and bilingual children with SLI have similar underlying deficits (Paradis et al., 2010a, 2010b), a greater

¹²³ In this context, language-dependency is viewed as “an abstract phonological property rather than a lexical or sub-lexical property” (Grimm & Hübner, forthcoming).

performance gap between TD and SLI is anticipated for both monolingual and bilingual children on the structurally more complex LD part of the LITMUS-QU-NWRT compared to the LI part of the task.

Several studies showed consistent effects of clinical group membership (i.e., SLI vs. TD) with no effects of bilingualism on performance in LITMUS-QU-NWR tasks, regardless of the procedure employed to determine the children's clinical status (e.g., clinical referral, L2-assessment, comprehensive L1/L2 assessment (see Grimm, 2022, for an overview). For example, Chilla et al. (2021) looked into the potential influence of L1-L2 typological differences on the performance of bilingual children with Arabic, Portuguese, and Turkish as L1 (age range: 5;6-9;0 yrs.) on the German and French versions of the LITMUS-QU-NWR task. They found that the task adequately identified SLI in bilinguals with no significant differences between the bilingual typically developing groups (regardless of their L1 and language dominance) and their monolingual peers on either part of the NWRT. It should, however, be noted that the diagnostic accuracy of the German version of the LITMUS QU-NWRT was slightly better than that of the French one; see also Tuller et al. (2018) for similar results. Scherger (2022) examined whether the German LITMUS-QU-NWR task is suited for identifying SLI in children who start acquiring German from the age of three onwards (i.e., early successive bilinguals) aged 6;3-8;6 years. The ROC curve analyses in Scherger¹²⁴'s study revealed excellent diagnostic accuracy for the global NWRT score (collapsing the LI and LD parts together) with sensitivity and specificity rates above 90% in both monolingual and bilingual groups.

Further studies looking into the diagnostic accuracy of the French and German LITMUS-QU-NWRTs (Abed Ibrahim & Fekete, 2019; Abed Ibrahim & Hamann, 2017; Chilla et al., 2021; de Almeida et al., 2017, 2019; dos Santos & Ferré, 2018; Ferré et al., 2015; Grimm, 2022; Grimm & Schulz, 2021; Grimm & Hübner, forthcoming; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018¹²⁵) found that the structurally more complex

¹²⁴ Note that Grimm (2022), Grimm and Schulz (2021) as well as Scherger (2022) used the short version of the German LITMUS-QU-NWRT. This version is based on Grimm and Hübner (forthcoming), who investigated the performance of 56 BiTD and BiSLI children aged between 8 and 10 years on the pilot long version of the German LITMUS-QU-NWRT task. The sample comprised simultaneous and early successive bilingual children with diverse L1s. Based on the discriminant analysis conducted in this study, a set of 36 test items emerged as those with higher discriminatory power.

¹²⁵ Note that de Almeida et al. (2017), Hamann and Abed Ibrahim (2017) as well as Tuller et al. (2018) will be discussed in detail in the introductory part of Chapter 5.

LD part of the NWRT did not disadvantage the BiTD children, who performed on par with their monolingual peers. On the contrary, compared to the LI part, the gap between SLI and TD was larger for the LD part leading to better diagnostic accuracy in both monolingual and bilingual populations. Furthermore, Ferré et al. (2015) and de Almeida et al. (2019) showed that the nonword test items with the alveolar lateral approximant [l] in word-internal coda¹²⁶ positions, as in [pil.fu], appeared to be sensitive to language impairment but not bilingualism. These results corroborate that phonological complexity is sensitive to phonological deficits not only in monolingual but also in bilingual children. Importantly, none of the previously mentioned studies found cumulative effects for bilingualism and SLI since no significant differences emerged between monolingual and bilingual children with SLI. Moreover, bilingual performance did not seem to vary by exposure measures, such as the age of onset of exposure and length of exposure to the L2 (Chilla et al., 2021; de Almeida et al., 2017; Tuller et al., 2018), making it a promising tool for the identification of language impairment in bilingual children, especially when used in combination with other diagnostic tools like sentence repetition tasks.

2.3.1.3 The parental questionnaire (PaBiQ)

Given that bilingual children vary considerably in patterns of their language exposure and use, which in turn influence the rate and outcome of language development (e.g., Chondrogianni & Marinis, 2011; Gathercole & Thomas, 2009; Grüter & Paradis, 2014; Hoff et al., 2012; Paradis, 2011, 2023; Paradis et al., 2010), results collected on tasks for bilingual subjects can only be adequately interpreted against background information about their language history. Having a clear idea about the child's early language development, family history of language difficulties, quality and quantity of L1/L2 input, and the family's socioeconomic status should help professionals determine whether a child's (poor) language performance is possibly linked to risk factors for language impairment (e.g., Restrepo, 1998) or to factors related to bilingualism. Risk factors for SLI include early language history, such as the age of emergence of first words or sentences, parental concerns about the child's language development, and familial history of language impairment. In contrast, factors related to bilingualism include timing (AoO) and the amount and quality of exposure to each of the bilingual child's languages.

¹²⁶ Note that word-internal codas were integrated only into the French version of the LITMUS-QU-NWRT as a complexity feature specific to French.

Numerous studies have used parental questionnaires to quantify bilingual experience in children and establish bilingual profiles; see Kaščelan et al. (2022) for a systematic review. In this thesis, the LITMUS repetition tasks were complemented with the LITMUS-Questionnaire for Parents of Bilingual Children (PaBiQ¹²⁷; Tuller, 2015). The PaBiQ incorporates questions from the ALEQ (Alberta Language Environment Questionnaire, Paradis, 2011) and the ALDeQ (Alberta Language and Development Questionnaire, Paradis et al., 2010) that were piloted in several studies and found to be best indicative of the presence of a developmental risk for SLI in child bilinguals, in addition to questions about exposure patterns to each of the child's languages. Thus, the questionnaire aims to gather information on risk factors for language impairment and document the quantity and quality of language input and use (past and current) within and outside the child's nuclear family. Based on the latter information, an experiential language dominance index can be calculated (see Unsworth, 2015; Unsworth et al., 2018). This serves as a baseline for adjusting the pathology cutoff criterion of monolingual norm-referenced tests according to the dominance status of the language being assessed (see Thordardottir, 2015a) and helps professionals better interpret the children's results on standardized tests, especially in cases where assessment in both of the child's languages is not possible (Boerma & Blom, 2017). The PaBiQ further provides a subjective estimate of the child's current L1 and L2 skills and delivers information on the parents' educational levels as a proxy for the family's socioeconomic status. Such information is crucial as language deficits associated with low SES are reported to negatively influence the performance on tasks with rich linguistic load, e.g., NWRTs and SRTs with word-like items (Chiat & Polišenská, 2016; Roy et al., 2014).

2.3.2 Estimating diagnostic accuracy of screening tools

A crucial indicator of the usefulness of an assessment tool is its diagnostic accuracy, i.e., the ability of a given test to discriminate between the existence and absence of a target condition. Diagnostic accuracy can be determined using different methods, such as sensitivity and specificity levels, likelihood ratios, predictive values, and the area under Receiver Operator Characteristic (ROC) curve; see Dollaghan (2007) and Šimundić

¹²⁷ In addition to the PaBiQ, the Bilingual Parents' Questionnaire (BIPAQ; Abutbul-Oz et al., 2012) was developed based on the ALEQ and ALDeQ questionnaires for clinical use in Israel, see Armon-Lotem (2018) for details.

(2009) for overviews. In language assessment, sensitivity refers to the proportion of children with language impairment identified as such within a subject pool (i.e., percentage of true positives), whereas specificity denotes the proportion of children with typical language development identified as such by the tests within the same subject pool, i.e., percentage of true negatives (Dollaghan & Horner, 2011; Oetting et al., 2008). Diagnostic accuracy represents the best tradeoff between sensitivity and specificity levels of a test (i.e., the highest possible sensitivity level alongside the highest possible specificity level). It is considered clinically informative if it reaches a level of 80% and above (Plante & Vance, 1994). In borderline cases, Plante and Vance (1994, p. 70) recommend opting for higher sensitivity rates rather than higher specificity rates, i.e., rather overidentifying a child as having language impairment than underdiagnosing her. This is because further differential diagnostics are likely to detect false positive cases and because the consequences of missing a genuine case of impairment are more far-reaching than the other way around.

The ROC analysis is frequently used to determine optimal sensitivity and sensitivity cutoff points (Dunn, 2014). In ROC analysis, the predictive accuracy of a test is estimated by the Area Under the Curve (AUC), whose value determines the test's overall accuracy. The values are interpreted as follows: an AUC of 1 indicates a perfect test; a value between .90 and 1 represents excellent accuracy; an AUC between .80 and .90 denotes good accuracy; an AUC between .70 and .80 signifies fair accuracy, while a test with an AUC ranging between .60 and .70 has poor accuracy and is factually worthless when AUC lies below .60 (see Swets et al., 2000).

A further method for determining diagnostic accuracy is using likelihood ratios (LRs), which are usually calculated based on previously established sensitivity and specificity levels. An advantage of LRs over other measures is that they are less likely to be influenced by variations in the properties of the test sample (Dollaghan & Horner, 2011). The positive likelihood ratio (LR+) indicates how likely it is that a score below a cutoff criterion to be present in language-impaired children and is calculated by the following formula: $[\text{sensitivity}/(1-\text{specificity})]$. Conversely, a negative likelihood ratio (LR-) is indicative of the likelihood of a score above a cutoff criterion to belong to a child without LI and is calculated as follows: $[(1-\text{sensitivity})/\text{specificity}]$. In tests with high diagnostic accuracy, i.e., tests with high sensitivity and specificity rates, the LR+ value is ≥ 10.0 , and the LR- value is < 0.1 . Tests with an LR+ ≥ 3.0 and an LR- value below 0.3

are considered to be clinically suggestive, while tests with an LR+ value under 3.0 and an LR– value > 0.3 are clinically uninformative (Dollaghan & Horner, 2011).

According to the Standards for the Reporting of Diagnostic Accuracy studies initiative (STARD), diagnostic accuracy represents “the best available method for establishing the presence or absence of a [a] target condition” against a reference standard (Bossuyt et al., 2015, p. 3). However, there is no agreed-upon standard for the identification of language impairment in bilingual children, and studies show significant variation in terms of sampling procedures, participant characteristics, and, most importantly, how reliably the clinical status of the children was established (Tuller et al., 2018, p. 3). Hence, it is crucial to try to avoid sources of diagnostic bias when evaluating the clinical utility of newly developed tools. Importantly, diagnostic accuracy should be calculated based on participant samples whose clinical status has been carefully verified. The latter issue was explicitly addressed in the BiLaD project, within which this thesis arose. As will become apparent in Chapter 3, special care was taken to recruit a relatively representative (diverse) bilingual sample and verify their clinical status. As anticipated, a significant proportion of the clinically referred bilingual children recruited for the project were reclassified as typically developing following the strict verification procedure outlined in **3.1.1** (see Hamann & Abed Ibrahim, 2017; Tuller et al., 2018 for details). As such, the diagnostic utility of the German LITMUS repetition tasks examined in this thesis can be determined with relative certainty.

3 Materials and Methods¹²⁸

The empirical studies in this dissertation were primarily conducted within the collaborative Franco-German research project BiLaD¹²⁹ (Bilingual Language Development: Typically Developing Children and Children with Specific Language Impairment), carried out at the University of Oldenburg, University of Bremen, and the University of Education Heidelberg in cooperation with the University of Tours in France. Using a similar methodology, the project investigated bilingual groups with Arabic, European Portuguese, and Turkish as first languages in Germany and France, which vary in their sociolinguistic and cultural settings, like community cohesion and L1 transmission (Condon & Régnard, 2010; Tucci & Groh-Samberg, 2008; Tuller et al., 2018). Furthermore, the typological differences between the children's L1s and the majority languages (German or French) allowed for cross-group comparisons, e.g., Abed Ibrahim et al. (2018), Chilla et al. (2021), Tuller et al. (2018), and Zebib et al. (2020b). Given that all studies follow the same research protocol, this chapter presents general methodology in order to avoid repetition across the subsequent empirical chapters. Nevertheless, each experimental chapter will have its own methodology section outlining relevant details for the respective study, such as participant details, specific test and statistical procedures, and any additions or alterations to the project's original experimental protocol.

This chapter is structured as follows. Section 3.1 provides detailed information on the participant recruitment procedure and how the children's clinical status as likely to be typically developing or likely to have SLI was determined by norm-referenced tests in the L1 and L2. Section 3.2 presents the standardized language tests used to assess the L1 and L2 in 3.2.1 and the standardized tests used to measure non-verbal IQ and working memory in 3.2.2. In 3.2.3, it proceeds to describe the LITMUS repetition tools and the LITMUS parental questionnaire. Finally, Section 3.3 provides general information about the data collection procedure and statistical treatment of the data. Note that a detailed

¹²⁸ Parts of this chapter are based on: Abed Ibrahim, Lina & Fekete, István (2019). What machine learning can tell us about the role of language dominance in the diagnostic accuracy of German LITMUS non-word and sentence repetition tasks. *Frontiers in Psychology*, 9, 2757. <https://doi.org/10.3389/fpsyg.2018.02757>

¹²⁹ The BiLaD project was financed by a joint grant (German DFG: HA 2335/6-1, CH 1112/2-1, and RO 923/3-1) and French ANR grant (ANR-12-FRAL-0014-01) to Laurice Tuller and her team.

description of the Arabic LITMUS sentence repetition task is separately provided in Chapter 4.

3.1 Participants

All studies presented in this thesis were conducted in line with the compliance form, transaction number 20120416505890730506, of the German Science Foundation and the recommendation of the “Kommission für Forschungsfolgenabschätzung und Ethik” (commission for the evaluation of research consequences and ethics) of the Carl-von-Ossietzky University of Oldenburg (rf. Drs. 21/16/2013). Parents or legal guardians of all participating minors provided written informed consent in their language of preference for both data collection and analyses. The research protocol was approved by the “Kommission für Forschungsfolgenabschätzung und Ethik” of the Carl-von-Ossietzky University of Oldenburg.

A large proportion of the participant pool for this thesis stems from the aforementioned BiLaD project, which targeted typically developing and language-impaired bilingual children with Arabic (Semitic Afroasiatic language), European Portuguese (Romance Indo-European language), and Turkish (Turkic Altaic language) as first languages. The latter L1s vary in their typological distance from German and are of particular relevance for the German context because a significant proportion of the immigrants residing in Germany are of Arab (Federal Statistical Office [Statistisches Bundesamt: Destatis], 2019), Portuguese (Hartman & Jackisch, 2013; Rinke & Flores, 2021) and Turkish origins (Akıncı et al., 2013; Chilla, 2022; Montanari et al., 2019). In addition to the bilingual groups, the project featured monolingual German-speaking children with and without language impairment as control groups. Data collection occurred from 2014 to 2017.

Monolingual and bilingual typically developing children were recruited from kindergartens, primary schools, and community-based schools offering instruction in the L1, e.g., weekend language schools. Children likely to have SLI, i.e., with a clinical diagnosis of SLI (ICD-10: F80.1/F80.2/F80.9 diagnoses, Bundesinstitut für Arzneimittel und Medizinprodukte, 2020), were recruited from specialized speech-language therapy facilities and kindergartens with special inclusion programs from different parts of Germany. As inclusion criteria for bilingual participants, children had to have at least 18

months of exposure to the L2 (German) and demonstrate functional bilingualism. Hence, children who failed to complete receptive subtests in the L1 were not included in the studies.

The age range of the participants was 5;6 to 9;0 years. This age range was chosen because it covers the last year of kindergarten and the first 2–3 years of primary school, where language assessment typically takes place. Moreover, in Germany, children usually start kindergarten between ages three and four, and bilingual children within this age range are likely to have already been exposed to the L2 for at least 18 months at the time of testing. In this case, “SLI can be determined with greater confidence [since] language growth trajectories [would have] stabilize[d]” (Tomblin, 2015, p. 536), and poor language performance is unlikely to be the result of insufficient exposure to the L2. This is crucial, since the overall goal of the BiLaD project and the studies presented in this thesis is to investigate the diagnostic accuracy of LITMUS repetition tools with participants who have a robust diagnosis of SLI.

Note that the number of participants in the studies included in this thesis varies depending on the state of recruitment when the studies were conducted. **Studies 1 and 3** (chapters 4 and 6, respectively) were supplemented by data from the follow-up research project BiliSAT¹³⁰ (Bilingual Language Development in School-age Children with/without Language Impairment with Arabic and Turkish as first languages) as well as by data collected by the doctoral candidate herself. The exact numbers and a detailed description of the participants featured in **Studies 1-3** are provided in the respective chapters.

3.1.1 Verification of clinical group membership

The initial classification of the recruited children as SLI or TD was based on clinical referral to a speech-language pathologist for intervention. However, given the high rates of over- and under-referral of bilingual children to speech-language therapy (Bedore & Peña, 2008; Grimm & Schulz, 2014), an extensive procedure based on standardized evaluation in each of the child’s languages in combination with information obtained via the PaBiQ (see 3.2.3.3) was applied to verify the children’s clinical status as $\pm$ language-

¹³⁰ The BiliSAT project was funded by DFG Grants to CH 1112/4-1S (to Chilla) and HA 2335/7-1C (to Hamann). It was carried out at the University of Oldenburg and the Europa-Universität Flensburg. Data collection occurred from 2017-2020.

impaired, see 3.2.1 for a detailed description. The clinical status of monolingual children was also verified to ensure that children likely to have SLI had true language impairment and not mere speech impairment, since the “[p]resence of speech problems is ... [a] salient factor in determining which children receive speech and language therapy services” (Bishop & McDonald, 2009, p. 603).

The verification process began with controlling for non-verbal intelligence using the German version of Raven’s Colored Progressive Matrices (CPM; Bulheller and Häcker, 2002). Only children with a non-verbal IQ score of 85 (the cutoff for ‘low average’ non-verbal intelligence) and above were included in the studies. As to assessment using formal tests, the commonly used reference standard for diagnosing SLI in monolinguals, according to which children are required to have a composite score below -1.25 z-score in at least two language domains (*EpiSLI* Model; Tomblin et al., 1996; see also Leonard, 2014a), was applied to the monolingual participants. As outlined in Hamann and Abed Ibrahim (2017), Tuller et al. (2018), and Chilla et al. (2021), the verification of the clinical status of the bilingual participants recruited for the BiLaD and BiliSAT projects was done in accordance with the recommendations of the COST Action IS0804 assessment committee, as summarized in Thordardottir (2015a, p. 343). Following these recommendations, the monolingual -1.25 SD cutoff criterion used by Tomblin et al. (1996) was adapted according to the dominance status of the language being assessed as estimated by the PaBiQ. Accordingly, a criterion of -1.5 SD was set if the child was being evaluated in her dominant language, whereas a cutoff of -2.25 SD was used for the child’s weaker language. For balanced bilinguals, the cutoff criterion was set at -1.75 SD for both languages.

Given that genuine language impairment would affect both languages of a bilingual child (Genesee et al., 2004), a bilingual child was assigned to the BiSLI group if she scored below language-dominance adjusted cutoffs in two language domains (on norm-referenced tests) in each of her languages. Five language areas relevant in this context were evaluated in each of the child’s languages (except for Turkish): phonology, morphosyntax comprehension, and production, as well as receptive and expressive vocabulary (Leonard, 2014a; Tomblin et al., 1996, 1997). Since expressive vocabulary is a notorious locus of difficulty for bilingual children, lexicon was counted as a single domain. A child was not considered impaired in this domain if expressive vocabulary was below the respective cutoff, but receptive vocabulary was above it.

In addition to standardized assessment, narrative samples were collected from each child in both of her languages using the picture materials provided by the LITMUS-Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2015). The collection of the narrative samples was done following the MAIN protocol (storytelling), where linguistically-trained native speakers of the respective languages conducted both collection and evaluation. It should, however, be noted that for the studies presented in this thesis, the narrative samples were not analyzed in terms of narrative macro- and microstructure but were instead used as spontaneous speech samples. Especially in borderline cases, the latter samples were consulted to gain an impression about the child's expressive language abilities in both of her languages and look for clinical markers for SLI, e.g., SVA errors, the use of infinitives, and verb placement errors in German (Clahsen, 1991a, 1991b; Hamann et al., 1998; Lindner, 2002; Rice et al., 1997).

Following the argumentation in Hamann and Abed Ibrahim (2017, p. 16) about problems encountered with standardized L1 tests in heritage contexts, and since the projects' general classification procedure does not isolate subgroups of SLI and might thus miss cases of selective impairment such as grammatical/syntactic, phonological or lexical SLI (cf. Friedmann & Novogrodsky, 2008), Hamann and Abed Ibrahim's (2017) modified "criteria for the identification of the bilingual clinical group" were applied to all studies presented in this thesis. Accordingly, a child was assigned to the BiSLI group if she had a selective impairment in the L2, i.e., if she performed below the dominance-adjusted cutoff in either morphosyntax or receptive vocabulary or phonology (not necessarily two domains in combination) and scored below norms in two domains in her L1 (one domain for Turkish) or showed poor performance of spontaneous production in both of her L1 and L2. It should be mentioned that the alternative procedure only concerned four children out of the entire sample; see 5.3.3 for particulars.

3.2 Materials

3.2.1 Standardized L1 and L2 tests

For the assessment of L1 and L2, we¹³¹ chose norm-referenced L1 and L2 tests frequently used by speech-language pathologists that cover the age range¹³² of our participants. For German, the LiSe-DaZ (Schulz & Tracy, 2011), which provides bilingual and monolingual norms, was selected for assessing morphosyntax. The short form of the WWT (Glück, 2011) was used to assess receptive and expressive vocabulary, and the screening version of the PLAKSS-II (Fox-Boyer, 2014) was used to evaluate phonology. We tried to assess the same language domains in Arabic, European Portuguese, and Turkish. For Arabic, we used the ELO-L test battery (Zebib et al., 2017, 2019), and for Portuguese, subtests of the PALPA-P (Castro et al., 2007) and GOLE (Sua-Kay & Santos, 2014) were used. For Turkish, we chose the TEDIL (Topbaş & Güven, 2011), which measures morphosyntactic comprehension and production as well as lexical semantics. The test, however, does not include a subtest for phonology. **Table 2** provides a detailed overview of the standardized language assessment tools.

¹³¹ “We” refers to the members involved in the projects BiLaD and BiliSAT.

¹³² In case of children older than the norming sample, the respective test authors were consulted regarding the possibility of norm-extension, e.g., LiSe-DaZ and ELO-L.

Table 2: Overview of standardized L1 and L2 tests

Language	Test	Language skill evaluated					Scoring method	Age range (norming sample)
		Phonology	Expressive vocabulary	Receptive vocabulary	Comprehension morphosyntax	Production morphosyntax		
Arabic	ELO-L ^a	Word repetition	Picture naming	Picture selection	Picture-sentence matching	Sentence completion	Individual subtest scores	3;0-7;11
German	WWT 6-10 ^b	--	Picture naming	Picture selection	--	--	Individual subtest scores	5;6-10;11
	LiSe-DaZ ^c	--	--	--	Constituent questions TVJT	Story, sentence completion, lead-in questions	Individual subtest scores	Monolinguals: 3;0-6;11 Bilinguals: 3;0-7;11
	PLAKSS-II ^d	Picture naming	--	--	--	--	Individual subtest scores	2;6-7;11
European Portuguese	PALA-P ^e	Nonword repetition	Picture naming	Picture selection	Picture selection	Sentence repetition	Individual subtest scores	5;0-9;11 (with missing norms for some age ranges for some subtests)
	GOL-E ^f	--	Antonyms naming	Word definition	--	Complex S from two simple S's	Individual subtest scores and global score	5;7-10;00
Turkish	TEDİL ^g	--	Picture naming	Picture selection	Picture selection	Sentence completion/ construction	2 composite scores, one for production, and one for comprehension	2;0-7;11

^aZebib et al. (2017, 2019); ^bGlück (2011); ^cSchulz and Tracy (2011); ^dFox-Boyer (2014); ^eCastro et al. (2007); ^fSua-Kay and Santos (2014); ^gTopbaş and Güven (2011);

TVJT=Truth value judgment task.

3.2.1.1 Standardized L1 tests

Standardized assessment of Arabic skills

For assessing oral proficiency in Arabic, we used the ELO-L comprehensive test battery (Evaluation du Langage Oral chez l'enfant Libanais – ‘Evaluation of Oral Language in Lebanese Children’, Zebib et al., 2017; see Zebib et al., 2019 for a detailed description), which is an adaptation of the French language test battery ELO (Evaluation du Langage Oral, Khomsi, 2001), a global assessment tool frequently used by SLPs in France. The ELO-L is the first norm-referenced language assessment battery for identifying language impairment in Lebanon and one of the very few available standardized assessment tools for Arabic considering all dialects (Zebib et al., 2019); see Shaalan (2017) and Rakhlin et al. (2021) for example tests for Gulf¹³³ Arabic. ELO-L’s normative sample consisted of 1718 Lebanese children aged 3 through 8 years raised in the kind of institutional bilingualism characteristic for Lebanon. The test exists in two versions: a long version for older children (6;0–7;11) and a short version for younger children (3;0–5;11), which is a subset of the longer version. Depending on the test form, administration lasts between 30 and 45 minutes. The test comprises five subtests assessing receptive vocabulary, expressive vocabulary, comprehension and production of morphosyntax, and phonology. All subtests begin with a training item to ensure that the child understands the task instructions.

The *Receptive Vocabulary* subtest is a picture selection task that includes 34 test items, all of which are nouns varying with regard to their estimated age of acquisition (Zebib et al., 2019). In this task, the child hears the target word and has to point at one of four pictures depicted on a PowerPoint slide. Each slide depicts the target stimulus in addition to three distractors (a phonological distractor, a semantic distractor, and an unrelated word). The subtask is usually scored by the total number of correct answers, where the child is awarded one point per correct response. The subtest also offers the possibility for further qualitative analyses, like error analysis by type of distractor. In this thesis, only the total number of correct responses is used for data analysis.

The *Expressive Vocabulary* subtest is a picture naming task. Here, the test administrator shows the child a picture on a PowerPoint slide and asks, “what is this?” or

¹³³ Gulf Arabic is an umbrella term used to refer to the closely related vernacular Arabic varieties spoken across the Arabic Peninsula and the Persian Gulf, e.g., Kuwaiti, Qatari and Saudi Arabic (Hole, 1995).

“what is he/she doing?”. The long version is composed of 70 test items (35 nouns and 35 verbs), while the short version consists of 54 items (27 nouns and 27 verbs). Similar to the lexical reception task, the items of this subtest vary in terms of age of acquisition of the words (Zebib et al., 2019). The original scoring method takes the peculiarities of the Arabic variety spoken in Lebanon into account and thus counts responses provided in French or English as correct. However, since the test was administered in the German context, only Arabic responses - regardless of whether colloquial or standard variety- were evaluated as correct, and each correct response was awarded 1 point.

Morphosyntax Comprehension is tested in a picture selection task that assesses comprehension of sentences varying in degree of morphosyntactic complexity and inferential load (Zebib et al., 2019). The short and long versions consist of 15 and 27 items, respectively. In this subtest, the child hears a sentence and is instructed to select the picture matching the stimulus sentence from an array of four pictures presented on a PowerPoint slide. Each correct response is awarded 1 point.

Morphosyntax Production is assessed via a sentence completion task. The subtest comprises 37 items (short version: 23 items, long version: 37 items) varying in degree of morphosyntactic complexity. The task is conducted as follows: the child is shown a picture and hears a sentence containing a certain morphosyntactic construction (the prime). Subsequently, the child is shown a second related picture and is requested to complete the target sentence, for which only the beginning is provided, “while taking into account the morphosyntactic transformations induced by the picture” (Zebib et al., 2019). The child can be awarded between one and three points for each completed sentence, depending on the number of required transformations.

The *Phonology* subtest is a word repetition task with 28 items varying in word length (1-4 syllables), syllabic complexity (absence or presence of consonant clusters), and familiarity. The test items with a lower degree of familiarity are characterized by their late age of acquisition and/or tend to contain very late-acquired phonemes, such as the uvular voiced stop /q/¹³⁴ as well as the emphatic voiceless alveo-dental stop /tˤ/ and fricative /sˤ/ phonemes (Dyson & Amayreh, 2000). According to the test authors, young children, in particular, are likely to process such items as non-words (Zebib et al., 2019). The task is scored by whole item accuracy (scoring scheme: 1 correct, 0 incorrect).

¹³⁴ This phoneme mainly occurs in Standard Arabic and has different allophonic realizations [ʔ, g, k] in the spoken varieties.

Genuine articulatory deficits and deviations arising due to inter-dialectal variations are not sanctioned.

For Lebanese Arabic, the ELO-L provides norms for each of the individual subtests. Moreover, separate norms exist for the production of nouns and verbs in the *Expressive Vocabulary* task. The z-scores are calculated by subtracting the obtained raw scores from the mean of the respective age group and dividing them by the standard deviation. In order to be classified as SLI based on the ELO-L, a child has to have at least two scores below language-dominance adjusted cutoffs, where, as mentioned in 3.1.1, only receptive vocabulary was considered for the verification of clinical group membership. The ELO-L was translated and adapted to a number of other varieties of Arabic by the test authors in collaboration with linguistically trained native speakers of the respective varieties (Algerian, Iraqi, Libyan, Moroccan, Palestinian/Jordanian, Syrian, and Tunisian). During the adaptation process, special attention was paid to vocabulary and the phonological characteristics of the different Arabic varieties, and oral stimuli were prerecorded for all of the adapted versions. Given that ELO-L was standardized and normed on the Lebanese population, z-scores of children assessed with the non-Lebanese versions must be considered cautiously.

Standardized assessment of European Portuguese

For assessing L1 abilities in European Portuguese (EP), we used subtests from the PALPA-P (Provas de Avaliação da Linguagem e da Afasia em Português; Castro et al., 2007), which is an adaptation of the *Psycholinguistic Assessments of Language Processing in Aphasia* (Kay et al., 1996). The full test battery comprises 60 different oral and written subtasks, is viewed as a linguistically well-constructed instrument for assessing EP as an L1 and is frequently employed in clinical settings. The test offers monolingual norms for children aged 5;0-9;0; however, the lexical subtasks lack norms for some of the age ranges under investigation. Hence, we used subtests from the PALPA-P to assess morphosyntactic production¹³⁵ and phonology and complemented them with subtests assessing receptive and expressive vocabulary from the GOL-E battery (Grelha de Observação da Linguagem – Nível Escolar; Sua Kay & Santos, 2014), which offers norms for children between 5;7 and 10;0 years of age and thus covers the entire age range

¹³⁵ Note that neither does PALPA-P nor GOL-E offer norms for morphosyntactic comprehension for the age range investigated here. As a result, only production of morphosyntax was considered for clinical classification based on the L1.

of children tested within the BiLaD project. The administration of the selected subtests takes about 25-30 minutes.

The phonology subtest (PALPA-P) is based on pseudoword repetition, where participants are requested to repeat 30 one-to-three-syllable pseudowords (10 items per syllable length) that resemble EP's existing words. Any segmental or prosodic alternation counts as an error unless it is attested as a potential dialectal variant within EP. In the morpho-syntax production task (PALPA-P), the child has to repeat 36 sentences verbatim. The stimuli are read aloud by a native speaker of EP and are grouped into three types with 12 sentences à type: active voice (progressive form), passive voice, and sentences with comparative constructions. The task is scored by correct identical repetition, where only articulatory and phonological errors are not penalized (for a detailed description, see Lein et al., 2017). Both tasks use binary scoring (correct = 1; incorrect = 0). The lexical reception and production tasks (GOL-E) consist of 10 items each, where children have to provide word definitions and antonyms. The latter task uses binary scoring, while in the word definition task, the child can be awarded up to two points per definition. For all tasks, the standard scores are established by comparing the raw scores to the standard values provided for the respective age range of the normative sample in the test manual and are later converted to z-scores. Similar to the ELO-L, two scores below dominance-adjusted cutoffs were required to classify a child as L1-SLI, where only receptive vocabulary was considered. Lein et al. (2016) note that the obtained z-scores appear to be very extreme in some cases for the PALPA-P subtasks due to the small size of the norming population (10-20 children per task) and are thus to be interpreted with caution.

Standardized assessment of Turkish skills

Turkish language abilities were assessed using the TEDİL (Türkçe Erken Dil Gelişim Testi, Topbaş & Güven, 2011) test battery, which is an adaptation of the English version of the Test of Early Language Development (TELD 3, Hresko et al., 1999) into Turkish. The test is frequently used by SLPs in Turkey as a screening tool for language delays and impairment. It was normed with a large sample of 1627 monolingual Turkish-speaking children (age range: 2;0-7;11) stemming from different regions in Turkey. The test takes 15-45 minutes to administer. It consists of a receptive and an expressive language subtest assessing lexical semantics, morphology, and syntax; however, it does not include a subtest for phonology. The test comes in two parallel forms, Form A and Form B, each

comprising a total of 76 items. In the studies reported in this thesis, Form A was exclusively administered to all participants for uniformity. Within the latter form, the receptive subtest contains 37 items: 24 semantic items and 13 morphology/syntax items, while the expressive subtest consists of 39 items: 22 semantic items and 17 morphology/syntax items. The test is conducted using a picture booklet, based on which the child is verbally instructed to provide a verbal or non-verbal response to prompts for each item, e.g., word-picture/sentence-picture matching, understanding relational vocabulary, sentence repetition, detecting ungrammaticalities via truth value judgment, sentence completion, and following instructions varying in complexity.

The TEDİL testing process does not necessarily begin with the first item. The entry point depends on the child's chronological age as calculated in years and months. This is “based on the assumption that the child can successfully perform all of the items that fall below the starting point” (Benner & Grim, 2012, p. 137). To determine the basal point, participants must provide three correct consecutive responses to items at their age level. For each correct response, the child is awarded one point. If the child fails three items in a row, testing proceeds backward until this criterion is met. Once the basal point is specified, testing continues forward, and the test is terminated when the child incorrectly responds to three successive items. The responses are documented in the respective answer sheet and later evaluated per the manual. The obtained raw scores are converted into age-equivalent standard scores (quotients) and z-scores using the conversion tables provided in the test's handbook.

It should be noted that the test does not provide separate norms for the individual linguistic subdomains assessed in the expressive and receptive subtests. Instead, composite standard scores exist for each of the comprehension and production subtests, collapsing morphosyntax and lexical semantics together. A composite total language score (combined expressive-receptive language score) can be calculated based on the standard scores in the latter two parts. Since TEDİL only offers a single production and a single comprehension score, encompassing two domains each, a child was classified as SLI based on the L1 if she scored below language-dominance adjusted cutoffs in either production or comprehension.

3.2.1.2 Standardized tests for assessment of L2-German

Test of vocabulary knowledge: Wortschatz und Wortfindungstest (WWT)

To assess the children's German vocabulary knowledge, the short version of the Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (WWT; Glück, 2011) was used, which is an expressive and receptive vocabulary measure standardized with a sample of 880 monolingual children (age range: 5;6-10;11) in Southern Germany (Bavaria). The test is available in three partially overlapping short forms. The choice of the form to be administered depends on the child's age. Each form consists of 4 training items and 40 test items: ten nouns, ten verbs, ten adjectives, and ten category nouns (hypernyms). In the expressive part, the participants are shown a picture and are requested to name the corresponding noun, action, antonym, or category noun. The receptive part targets the exact same items as the expressive part in the form of a picture selection task, where the child hears a word and has to select a picture out of an array of four. The receptive part can either be conducted right after the expressive part or in a separate session. In our case, the receptive subtest immediately followed the expressive part. For this reason, only those items answered incorrectly in the expressive part were evaluated in the receptive part (following the test manual). Items with correct responses in the expressive part were viewed as correct in the subsequent receptive part and were awarded a point. Based on the test manual's norm tables, the obtained raw points were converted into standard scores (t-scores and percentile ranks), which were subsequently transformed into z-scores. Since the test does not offer bilingual norms, monolingual cutoff points for identifying language impairment were adapted according to the children's language dominance following Thordardottir (2015a). As mentioned in 3.1.1, lexicon was viewed as one domain, and only receptive vocabulary was considered to verify the children's clinical status.

Test of morphosyntactic knowledge: LiSe-DaZ

Children's morphosyntactic abilities were assessed using the standardized test battery LiSe-DaZ (Linguistische Sprachstandserhebung – Deutsch als Zweitsprache; Schulz & Tracy, 2011), which is the only available standardized German test with separate norms for monolingual and early successive bilingual children with respect to their length of exposure. The norming sample included 303 monolinguals (age range: 3;0-6;11) and 609 early successive bilingual children (age range: 3;0-7;11) from 8 different German states (Schulz & Tracy, 2011, p. 86f).

The test consists of a comprehension and a production module. The comprehension module features three subtests: a subtest assessing the comprehension of telicity (verb semantics, see Penner et al., 2003) in the form of a TVJT, a subtest assessing constituent (argument and adjunct) *wh*-questions (syntax/semantics) using a picture booklet, and a subtest evaluating sentential negation (syntax/semantics, see Wotjecka et al., 2011) in the form of TJVT. The production module is elicited via a picture story task. It consists of 8 subscales, three of which examine aspects that represent core areas of German morphosyntax (Schulz & Grimm, 2019), namely subject-verb agreement (SVA), case marking, and sentential complexity (*Entwicklungsstufe*: developmental level). The subscale sentential complexity measures the mastery of complex constructions on a scale from 1 to 4, where 1 corresponds to single-word utterances and 4 to the production of embedded clauses. The child's developmental level corresponds to the highest level at which the child produced at least three utterances. The remaining subscales assess the following word classes: prepositions, focus particles, main verbs, auxiliary verbs, and subordinations (morphosyntax-lexicon). All tasks except for sentence complexity and SVA provide T-values and can be used to identify language impairment in monolingual and early successive bilingual children (see Lemmer, 2018; Schulz & Schwarze, 2017; Wotjecka, 2020 for particulars). A limitation of the test is that it does not offer norms for simultaneous bilingual children or bilingual children older than eight years.

The assessment procedure follows Grimm and Schulz (2014) with a minor modification, namely that a child was assigned to the language-impaired group if she scored 1.25 SD (instead of the 1 SD recommended by the test authors) below $T = 50$ in two out of the nine subscales providing T-values for early successive bilingual learners. The -1.25 SD cutoff was chosen because it is recommended by Tomblin et al. (1997) and is frequently used in clinical studies (e.g., Eadie et al., 2014). For the monolingual children, the available monolingual norms were applied. Sequential bilinguals with an AoO ≥ 24 months were compared to the bilingual norms¹³⁶. As to simultaneous bilinguals with an AoO < 24 months, Schulz and Tracy (2011) recommend using monolingual or bilingual norms. Consequently, we used monolingual norms if a simultaneous bilingual¹³⁷ child was German-dominant or balanced according to the PaBiQ. On the other hand,

¹³⁶ Since the LiSe-DaZ provided bilingual norms, there was no need to apply the adjustments recommended by Thordardottir (2015a), which are mainly applicable mainly to 2L1 children.

¹³⁷ Monolingual norms were applied to German dominant or balanced 2L1 children, since they were expected to rather behave like monolingual German children.

bilingual norms were applied for simultaneous bilinguals who were L1-dominant. For bilingual children older than eight years, a child is identified as having SLI if she scores 0.5 SD below $T = 50$ in at least two subtests that provide t -values, following Grimm and Schulz's (2014) suggestion for norm extensions.

Screening tool for phonological development: PLAKSS-II

Children's phonological abilities in German were evaluated using the screening version¹³⁸ of the PLAKSS-II (Psycholinguistische Analyse kindlicher Sprechstörungen, Fox-Boyer, 2014). The test is based on the classification model of Dodd (1995/2005) and is frequently used by German SLPs and health practitioners as a screening tool for speech sound disorders. The stimuli of the screening material were carefully chosen to cover the full range of German phonemes (every phoneme occurs at least once) and include consonant clusters characteristic of German.

The test consists of 30 items and is administered as a picture-naming task. If the child does not recognize or fails to provide the targeted word, the test administrator assists the child by providing a context sentence that would elicit the targeted word. For example, if the target word is *Elefant*, the experimenter says: *Der Rüssel gehört dem.....* 'the trunk belongs to the.....'. If the child still does not provide the required response, the experimenter can read the word aloud and ask the child to repeat it. The participants' responses were audio-recorded, phonetically transcribed, and verified by linguistically trained research assistants. The latter phonetic transcripts served as the basis for subsequent analyses of articulatory and phonological processes. Depending on the observed processes, children's speech development status was classified into two main categories using a binary scheme: 0 = developmentally age-appropriate phonological development, 1 = phonological delay or pathological development. Children with the status '1' were viewed as impaired in the domain of phonology.

¹³⁸ PLAKSS-II also exists in a long form (99 test items) with normative data for monolingual children aged 1;6-6;0 (Fox-Boyer, 2014). It should be noted that the short screening version is only to be used as a first estimate of phonological abilities and does not suffice for reliable differential diagnostics. For the latter purpose, it is recommended to use the long version of the test.

3.2.2 Standardized tests for assessment of non-verbal intelligence and working memory

Assessment of non-verbal IQ

The German version of Raven's Colored Progressive Matrices (CPM; Bulheller & Häcker, 2002) was administered in order to obtain a rough estimate of the children's non-verbal intelligence. The test consists of 36 matrix reasoning tasks, in which the child is requested to solve a visual pattern problem by choosing the correct missing pattern piece from an array of six. The test is particularly suited for assessing fluid intelligence in children with limited expressive language abilities since it requires minimal verbal instructions to the child. Verbal responses are not required, and the child can respond by merely pointing to the corresponding pattern piece. In order not to disadvantage bilingual children, the test instructions were given in the child's dominant language.

Working memory measures

Verbal short-term memory (VSTM) and verbal working memory (VWM) were assessed through the forward and backward digit span subtests of the German version of the Wechsler Intelligence Scale for Children (WISC-IV Petermann & Petermann, 2011). Compared to other serial-recall tasks, such as non-word repetition, tasks involving digits are assumed to be less sensitive to lexicality effects and differential effects of language exposure (Gathercole et al., 1999; Majerus et al., 2006; Messer, 2010) since number knowledge is acquired relatively early (Gathercole & Adams, 1994) and is expected to be well-entrenched by school entry (Engel de Abreu et al., 2013). For this reason, digit recall tasks are viewed as a fair measure for assessing working memory in bilingual children, especially since they draw on item information sampled from a closed pool and are very likely to be familiar even to children with less experience with the language of testing (Engel de Abreu et al., 2013).

Forward Digit Span (FDS) was employed to measure VSTM (phonological loop). In this task, the child is aurally presented with a series of digits and is asked to repeat them in the same order they were presented. The task consists of eight lists of digits progressively increasing in length from 2 to 9 digits. For each sequence length (span), the child completes two test trials and is awarded 1 point per correct trial (maximum 16 points for total raw score). Any alternation to the presented sequence (wrong order, substitution, addition, or omission of digits) counts as a failed attempt. The task is discontinued if the

child incorrectly repeats both sequences of the same length. The highest attainable forward digit span score is 9, corresponding to the number of digits in the longest list in which the child succeeds in at least one of the two test trials.

Backward Digit Span (BDS) was used to assess VWM. In this task, children are instructed to repeat sequences of digits in reverse order. Such complex-span tasks involve both storage and simultaneous processing since the participant is not only required to temporarily store the incoming stimuli in the phonological loop but also manipulate them, resulting in additional processing demands (see Gathercole et al., 2004a; La Pointe & Engle, 1990). The task comprises eight lists of digits (2 test trials per list), which progressively increase in length from 2 to 8 digits. To familiarize the child with the task, practice items are offered with two-digit sequences to ensure that the child understands the concept of “backward recall”. Analog to FDS, every correct response is awarded 1 point. The test is aborted if the child fails two consecutive trials within a list. The backward span procedure resembles that of the FDS, i.e., the VWM span corresponds to the longest successfully repeated sequence. In both the FDS and BDS subtests, numbers were read at the rate of one digit per second.

3.2.3 The LITMUS tools

3.2.3.1 The German LITMUS Sentence Repetition Task

The original long version¹³⁹ of the German LITMUS sentence repetition task, developed by Hamann et al. (2013), was constructed in close parallel to the French LITMUS-SRT (de Almeida et al., 2017; Fleckstein et al., 2018) and contained 64 sentences. The task version used in this thesis comprises 45 sentences divided into three levels of increasing syntactic complexity. Each level consists of 5 conditions with three items per condition. In terms of length, sentences range from 7 to 14 syllables, controlling for the number of syllables within each level. The degree of an item’s structural complexity depends on the presence of complexity factors, such as *wh*-movement, clausal embedding, and intervention or combinations thereof, where the latter may add difficulty to the presence of two propositions. The sentence stimuli were constructed from lexical items presumed familiar to children through kindergarten storytelling (e.g., Viking, pirate, witch, king,

¹³⁹ The long version of the German LITMUS-SRT was shortened in order to accommodate the needs of the age range of the participants recruited for the BiLaD project (see Hamann et al., 2017 for particulars).

prince, princess, knight), although some items turned out to be slightly phonologically challenging; item frequency was controlled throughout the task.

As can be seen in **Table 3** and the examples below, Level 1 consists of simple declaratives (7-9 syllables) and focuses on Subject-Verb-Agreement (SVA), tense, and the sentence bracket (see (1)). Level 2 (9-13 syllables) includes two types of object questions: non-referential bare *wh*-questions with overt accusative case marking on the *wh*-operator (Wen ‘who-masc.-acc.’), and referential (discourse-linked) object *wh*-questions (*which* NP-questions)¹⁴⁰ with overt accusative case and gender marking on the *wh*-operator (*Welchen* ‘which-masc.-acc.’) followed by an intervening lexical noun phrase, see (2a) and (2b). Level 2 further contains non-finite and finite (see (3)) complement clauses. Finite complement clauses contrast with coordinate structures, which serve as control items (two propositions but no embedding). Level 3 (11-12 syllables) comprises the most complex constructions and tests long passives, topicalizations (see (4)) as manifestations of the V2-property¹⁴¹ of German, subject relative clauses, and object relative clauses without (i.e., with a pronoun subject) and with intervening lexically restricted DPs, see (5a) and (5b).

Note that although German has morphological case marking on elements in agreement dependencies with the governing noun, only masculine singular nouns, such as *der Clown*, require differentiating between nominative and accusative case marking on determiners, interrogative and relative pronouns, as in examples 2a/2b and 5a/5b. **Table 3** gives an overview of the test conditions. For more details on the construction of the German LITMUS-SRT, see Hamann et al. (2017) and Hamann and Abed Ibrahim (2017). A complete list of the test items is provided in the Appendix (see I).

¹⁴⁰ Which-NP questions are considered to be structurally and semantically more complex than bare *wh*-questions due to the presence of a lexically restricted subject DP that intervenes between the displaced constituent (the object) and its trace and their computation involves the partitioning of a subset of a set (Avrutin, 2000; Ebbels & van der Lely, 2001; Friedmann & Novogrodsky, 2011; Guasti et al., 2012; Roesch & Chondrogianni, 2016; Thompson et al., 1999; but see Biran and Ruigendijk (2015) for a discussion on the interaction between intervention and object case marking in German).

¹⁴¹ German is a verb-second (V2) language, which requires the finite verb (marked for tense and agreement) to be the second constituent of main (root) clauses (Haider, 1997). Like other Germanic languages (except for English), it allows different types of constituents (e.g., complements, adverbials) to occupy sentence initial position in main clauses (e.g., topicalization), where only one constituent can linearly precede the finite verb (cf. Chomsky, 1986; Eisenberg, 1999, 2013).

Table 3: German LITMUS-SRT: an overview of test conditions

Level 1	Level 2	Level 3
SVO-Present	object who-question (bare <i>wh</i>)	long passive
SVO-Simple past	object which-question (<i>wh</i> -NP)	topicalization (OVS)
sentence bracket (Aux)	coordination	subject relative clause
sentence bracket (Particle verbs)	non-finite complement clause	object relative clause without intervener
„Werden“ control	finite complement clause	object relative clause with intervener

(1) Sentence bracket

Die	Clowns	haben	den	Hund	gesucht
The/ <i>nom/pl</i>	clowns	have	the/ <i>acc-m</i>	dog	searched for

‘The clowns searched for the dog’

(2a) Object who-question: Bare *wh*

Wen	untersucht	der	Tierarzt	gerade?
Who/ <i>acc-m</i>	examines	the/ <i>nom-m</i>	veterinarian	now?

‘Who(m) is the veterinarian examining now?’

(2b) Object which-question: *wh*-NP

Welchen	Jungen	mag	der	Zauberer?
Which/ <i>acc-m</i>	boy	likes	the/ <i>nom-m</i>	magician?

‘Which boy does the magician like?’

(3) Finite complement clause

Der	Clown	möchte,	dass	die	Frau	den	Mann	besucht.
The/ <i>nom-m</i>	clown	wants,	that	the/ <i>nom-f</i>	woman	the/ <i>acc-m</i>	man	visits

‘The clown wants that the woman visits the man’

(4) Topicalization

Den	Clown	umarmt	der	Wikinger	am liebsten
The/ <i>acc-m</i>	clown	hugs	the/ <i>nom-m</i>	viking	best

‘The clown, the Viking hugs best’

(5a) Object relative without intervention

Der	Roboter	sieht	den	Mann, den	er	malen	will	
The/ <i>nom-m</i>	robot	sees	the/ <i>acc-m</i>	man	who/ <i>acc-m</i>	he	draw	wants

‘The robot sees the man who(m) he wants to draw’

(5b) Object relative with intervention

Ich	sehe	den	Roboter, den	der	Bäcker	malt.	
I	see	the/ <i>acc-m</i>	robot	who/ <i>acc-m</i>	the/ <i>nom-m</i>	baker	draws

‘I see the robot who(m) the baker draws’

The test stimuli were prerecorded in a soundproof booth by a female native speaker of German. Subsequently, they were equated for the intensity of pitch and amplitude using PRAAT (Boersma & Weenink, 2013), pseudorandomized, and integrated into a child-friendly PowerPoint Presentation. The task starts with two warm-up items and takes about 10 minutes to administer with no abort criterion. At the end of each row (9 sentences), a friendly sun appears as a form of positive feedback encouraging the child to continue the task (see **Figure 2**).

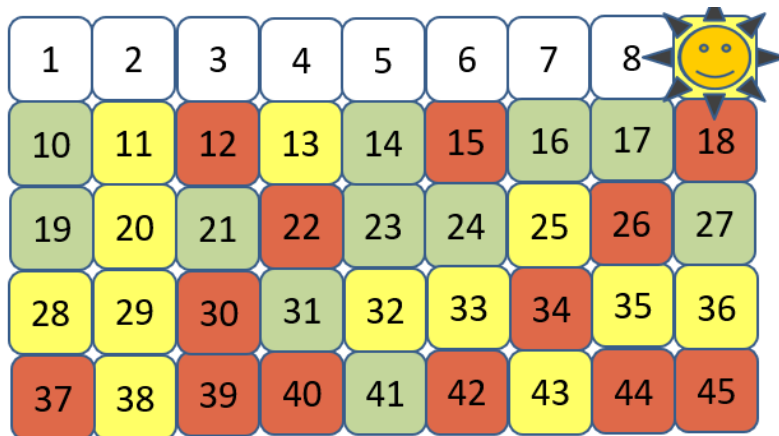


Figure 2: PowerPoint Slide from the German LITMUS-SRT (first row)

Each test item is played only once unless there is a disruptive factor causing the child not to hear the stimulus sentence. The test items can be scored in two different ways: identical repetition (SRT_Id), i.e., whole item accuracy, which only disregards phonological errors, and correct target structure (SRT_Tar), which measures whether a particular structure has been mastered in principle (see Marinis & Armon-Lotem, 2015, for scoring measures). Scoring by identical repetition is the fastest and simplest way of scoring; however, difficulties that bilingual children encounter with vocabulary and case/gender marking could surface using this scoring method and penalize them. Scoring by correct target structure compensates for typical L2 errors by abstracting away from lexical substitutions, omissions of attributive adjectives, and systematic recurrent case as well as gender errors, as long as they do not affect the realization of the targeted sentence structure, see examples (6) to (8). Comparing these scoring methods has shown that the latter scoring method improves the diagnostic accuracy of the German LITMUS-SRT (see Abed Ibrahim & Fekete, 2019; Hamann & Abed Ibrahim, 2017; Hamann et al., 2017). The TOLD-P-4 (Newcomer & Hammill, 2008) binary scoring scheme was chosen, which allocates a score of 1 if the sentence was repeated correctly and a score of 0 if not, regardless of the number of errors. Clinicians favor this scoring scheme since it is faster and less susceptible to scoring errors compared to the fine-grained CELF (0-3) scoring system (Semel et al., 1995), which allocates a different score depending on the number of errors in the utterance.

(6) Target structure: sentence bracket

Der	Prinz	hat	die	Prinzessin	umarmt
The/ <i>nom-m</i>	prince	has	the/ <i>acc-f</i>	princess	hugged

‘The prince hugged the princess’

Child repetition:

Der	König	hat	die	Prinzessin	umarmt
The/ <i>nom-m</i>	king	has	the/ <i>acc-f</i>	princess	hugged

Lexical substitution error, score for correct target structure: 1

(7) Target structure: long passive

Der	Clown	wird	von dem	alten	Roboter	gefilmt
The/ <i>nom-m</i>	clown	is	by the/ <i>dat-m</i>	old	robot	filmed

‘The clown is filmed by the old robot’

Child repetition:

Der	Clown	wird	von den	alten	Roboter	gefilmt
The/ <i>nom-m</i>	clown	is	by the/<i>acc-m</i>	old	robot	filmed

Systematic case error in passives (also recurring in monolingual data): dative case marking of the determiner in the prepositional phrase ‘von dem’ is replaced by accusative case marking ‘von den’, score for correct target structure: 1

(8) Target structure: long passive

Die	nette	Ärztin	wird	von der	Mutter	umarmt
The/ <i>fem-m</i>	nice	doctor	is	by the/ <i>dat-f</i>	mother	hugged

‘The nice doctor is hugged by the mother’

Child repetition:

Die	Ärztin	wird	von der	Mutter	umarmt
The/ <i>fem-m</i>	doctor	is	by the/ <i>dat-f</i>	mother	hugged

‘The doctor is hugged by the mother’

Lexical error: omission of the attributive adjective ‘nice’, score for correct target structure: 1

(9) Target structure: Subject-Verb-Agreement, past, 3P/sg.

Der	Clown	umarmte	die	Köchin
The/ <i>nom-m.</i>	clown	hugged	the/ <i>acc-f</i>	cook

‘The clown hugged the cook’

Child repetition:

Der	Clown	umarmte	den	Köchin
The/ <i>nom-m</i>	clown	hugged	the/<i>acc-m</i>	cook

Gender error with overt accusative case marking on masc./sg. determiner, score for correct target structure: 1

3.2.3.2 The German LITMUS-QU-Nonword Repetition Task (LITMUS-QU-NWRT)

The long version of the German LITMUS-QU-NWRT (Grimm et al., 2014) employed in this thesis is composed of two parts (subtests): a structurally less complex quasi-universal (i.e., language-independent) part (LI) and a language-specific, i.e., language-dependent part (LD), which incorporates more complex structural aspects. The items of both parts were constructed to be maximally distinct from existing German words and range in length from one to a maximum of three syllables to reduce phonological working memory load. In both subtests, “a trochaic ‘CVCV’¹⁴² shape forms the base varying systematically in complexity” (Grimm & Schulz, 2021, p.269). Throughout the whole task, phonological complexity is systematically varied at the segmental (consonantal), syllabic (presence of branching onsets or coda), sequential (position of the cluster within the nonword), or word levels (see dos Santos & Ferré, 2018; Grimm & Hübner, forthcoming; Grimm & Schulz, 2021 for details). At the word level, phonological complexity is realized by attaching an extrametrical syllable to the right of the trochaic foot, i.e., an unfooted syllable that “cannot be integrated into [the] trochaic foot [...] and is adjoined to the prosodic word in German” (Grimm & Hübner, forthcoming; Wiese, 1996). The structural representation of the prosodic word and foot of an example trisyllabic non-word is depicted in **Figure 3** (see also Féry, 2008). In this example, the nonword /'pufaki/ is formed by right-adjointing the unfooted syllable /ki/ to the disyllabic prosodic foot /'pufa/.

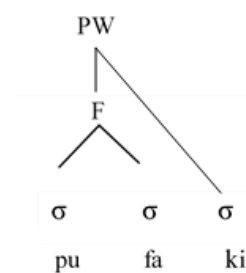


Figure 3: Representation of word-level complexity in a trisyllabic nonword (adapted from Grimm & Hübner, forthcoming). Abbreviations: PW: prosodic word, F: foot, σ : syllable

The language-independent part consists of 23 test items and seven control items to ensure that the children can produce the individual segments when they are not part of a cluster, e.g., /kip/, /'pilu/. All items of the LI part were constructed using phonemes, and phonotactic constraints attested in the vast majority of the world’s languages

¹⁴² ‘C’ stands for consonant, ‘V’ for vowel and “’” marks the stressed syllable.

(Maddieson et al., 2011), i.e., phonemes that are “compatible with cross-linguistically diverse constraints on lexical phonology” (Chiat, 2015, p. 138). The phonemes used in the LI part are (/a, i, u/ and /p, k, f, l/). Unlike the nonwords of the LITMUS-CL-NWR task discussed in Chiat and Polišenská (2016), the nonwords of the German LITMUS-QU-NWR task are shorter and are not only composed of simple CV sequences but also include syllables with initial consonant clusters ‘#CCV’ or closed syllables of the type ‘CVC#’, which are typologically well-attested, albeit their relative complexity (Maddieson, 2006).

The language-dependent part contains 36 items (32 test items and four control items), adhering to the same construction principles of the LI part. It employs the same vowels and consonants used in the LI part, adding a combination of the coronal fricatives /s, ʃ/ and an obstruent (henceforth ‘#sO’) in word-initial (e.g., ‘sCVCVC in /skifup/) and word-final positions (e.g., ‘CVCVCVCs in /kapifaps/) as a complexity feature specific to German (and some other languages, e.g., English and Dutch). Such sC sequences violate the Sonority Sequencing Principle and are considered to be phonologically more complex than other types of onset clusters (Selkirk, 1984). Moreover, they appear cross-linguistically less frequently than other types of clusters (Goad & Rose, 2004). Note that the voiceless alveolar fricative /s/ is the most frequent consonant cross-linguistically (Maddieson, 1984). It is mastered at a relatively early age by monolingual German-speaking children (at around age 2;11), while the post-alveolar fricative /ʃ/ is acquired at age 4;11 (Fox & Dodd, 1999). Hence, the latter two segments are not expected to pose difficulties for the participating children given their older age. In line with Wiese (1991, 1996), the authors of the test assume that “/s/+Obstruent is represented as a complex segment in word-initial position ... and as an extrasyllabic consonant in word-final position” in German (Grimm & Schulz, 2021, p. 270). In this case, the /s/ is attached to the syllable if it occurs word-initially and to the Prosodic Word if it occurs word-finally. Note that while many languages, including typical migrant languages¹⁴³, appear to allow #sO sequences at the surface level, i.e., phonetic representation, they do not necessarily share the same phonological representation. For example, in English (Hammond, 2003) and German (Wiese, 1996), #sO clusters comprise branching onsets, whereas the /s/ is

¹⁴³ Among the languages investigated in the BiLaD project, Turkish (van der Hulst & van de Wijer, 1991) and European Portuguese (Mateus & d’Andrade Pardo, 2002) allow #sO clusters word-initially at the phonetic level, whereas such sequences are not attested in all Arabic dialects or only occur in loanwords (Watson, 2007), see Grimm and Schulz (2021) for a discussion.

attached to a preceding empty nucleus in other languages like Polish (Gusmann, 2007; Rubach, 1997) and Portuguese (Mateus & d'Andrade Pardo, 2002), see also Hall (2001) for a discussion on the extrametricality of /s/ at word-initial and word-final syllable margins. Hence, difficulties in producing an #sO cluster cannot only be attributed to the absence of this type of cluster in the children's L1 but can also result from varying phonological representations (Fikkert & Freitas, 2004; Grimm & Schulz, 2021). Given its construction, the LD part is considered structurally more complex than the LI part. At the same time, it is less dependent on language-specific knowledge when compared to more traditional language-specific NWRs (e.g., Rispens and Baker, 2012) that draw on the entire phoneme inventory and include many more properties specific to the target language (Chiat, 2015; Chiat & Polišenská, 2016). An overview of segments and syllable types is given in **Table 4**. For a complete list of the test items, see **Appendix II**. Task administration takes about five minutes, and the nonwords are presented to the child in a pseudo-randomized order via an animated PowerPoint Presentation. The test items were presented via noise-canceling headphones to avoid external disruptions and ensure a consistently good audio output quality. At the beginning of the task, children were told that an alien from another planet would appear on the screen and try to teach them his language (format adapted from Engel de Abreu et al., 2013); see **Figure 4**.

In line with Chiat and Polišenská (2016), the test was scored by whole item accuracy (percentage of items correct). This scoring method is better suited for clinical purposes and is shown to be informative (Boerma et al., 2015; Roy & Chiat, 2014). A response was rated as correct if all consonants and vowels and their sequencing corresponded to the target form. Phoneme omissions, substitutions, or additions¹⁴⁴ are regarded as incorrect. Non-responses were rated as 'incorrect'. Since the task mainly targets bilingual children, L2 errors such as voicing of consonants (/pilu/ → /bilu/) or slight vowel alternations (/faku/ → /fako/) are disregarded. Furthermore, the substitution of /f/ through [s] or an interdental pronunciation of /s/ are not counted as errors since this does not result in a phonemic contrast when they occur as part of an #sO or #Os cluster in German (Grimm & Hübner, forthcoming).

¹⁴⁴ Note that some studies such as Boerma et al. (2015) and Dollaghan and Campbell (1998) did not penalize additions of phonemes based on the argument that additions do not result in loss of phonological information. However, addition errors can reflect imprecise preservation of the phonological input and are thus treated as errors in many NWR tests (Chiat & Polišenská, 2016; Sorenson Duncan & Paradis, 2016; Wagner et al., 1999). For example, Burke and Coady (2015) found that English-speaking monolingual children with SLI showed significantly higher phoneme addition rates than their typically developing peers.

Table 4: German LITMUS-QU-NWRT: an overview of segments and syllable types

	vowels	consonants	syllable types	examples
Language-Independent part (LI) 30 items 23 test items 7 controls (e.g., faku, paf)	/a, i, u/	/p, k, f, l/	CV CCV CVC#	faku flaplu kafip
Language-Dependent part (LD) 36 items 32 test items 4 controls (e.g., kif, sapi)	/a, i, u/	/p, k, f, l/ plus /s/ /ʃ/	same syllable types plus #sCV #sCCV Cs# internal /s/	skupifla Splaklu fikapuks kufiski



Figure 4: PowerPoint Slide from the German LITMUS-QU-NWRT

3.2.3.3 The Questionnaire for parents of bilingual children LITMUS-PaBiQ

The Parents of Bilingual Children Questionnaire (PaBiQ; Tuller, 2015) was used to gather relevant background information about possible risk factors for language impairment as well as the quantity and quality of language input/output in the child's environment. It contains parts of the Alberta Language and Development Questionnaire (ALDeQ, Paradis et al., 2010), with focus on early developmental milestones, current L1 abilities, behavior patterns, and activity preferences as well as family history, in addition to parts of the

Alberta Language Environment Questionnaire (ALEQ, Paradis, 2011), with focus on the child's linguistic environment and current language input and use. The PaBiQ was translated into several languages, including Arabic, Turkish, European Portuguese, and German (see **Appendix III** for the German version). Several indices can be calculated based on information obtained via the PaBiQ, such as the No-Risk Index and the Language Dominance Index (LDI).

The No-Risk Index

The PaBiQ incorporates questions about developmental risk factors for SLI, which are synthesized into a global No-Risk Index¹⁴⁵, for which a maximum of 23 points can be attained. This index is arrived at by collapsing the scores of the Positive Early Development Index, which is associated with the timing of early language developmental milestones, and the Family History Index, which is associated with the presence of oral/written language disabilities in the family. The Positive Early Development Index (/14 pts) is calculated by adding up the sub-scores for the age of first words (≤ 15 mo. = 6 pts; 16-24 mo. = 4 pts; > 25 mo. = 0 pts), age of first multiword utterances (≤ 24 mo. = 6 pts; 25-30 mo. = 4 pts; > 31 mo. = 0 pts), and early parental concerns (yes = 0 points; no = 2 points). The familiar risk for SLI (/9 pts) is indexed by the existence of first-degree relatives (mother, father, siblings) with reading/spelling problems, difficulties understanding others when they speak, or difficulties expressing themselves orally. Children with a negative family history of language problems are awarded a maximum of 9 points (3x3: 1 point per family member per type of language difficulty).

Boerma and Blom (2017) investigated the influence of language impairment (LI) and bilingualism on the latter two indices and looked into their diagnostic accuracy. In line with Paradis et al. (2010), they reported strong adverse effects of LI on Early Language Development and showed that it was a strong predictor of language impairment in both monolingual and bilingual children confirming previous findings that late onset of first words and sentences in at least one language is a risk factor for SLI (cf. De Houwer, 2009; Reilly et al., 2010). Concerning the Family History Index, Boerma and Blom (2017) observed an adverse effect for language impairment in the monolingual group but not in the bilingual one and concluded that, due to cultural factors, "Family

¹⁴⁵ The higher the value of the No Risk Index, the less likely the child is to be at risk of having SLI. The same applies to the index of Positive Early Development.

History as reported by parents may [...] be less reliable as an index of LI in bilingual children than in monolingual children” (p. 73). The Positive Early Development Index also yielded promising diagnostic results in the study by Tuller et al. (2018), who found it to be the leading factor explaining performance differences between BiSLI and BiTD in German and French LITMUS nonword and sentence repetition tasks. In light of previous findings and since the Positive Early Development Index is a component of the No-Risk Index, the former was chosen for statistical analyses, especially since using both indices could result in serious collinearity issues.

The Language Dominance Index

The PaBiQ further allows the calculation of a purely experiential Language Dominance Index (LDI) as a differential score between the L1 Exposure Index (the relative amount of exposure to the L1) and the L2 Exposure Index (the relative amount of exposure to the L2, i.e., German). For each of the child’s languages, a total of 50 exposure/use¹⁴⁶ points could be attained using the German PaBiQ¹⁴⁷. The Exposure Index is calculated for each of the child’s languages separately based on AoO (/ 4 pts), LoE¹⁴⁸ (/4 pts), frequency of early (before the age of four) language use and exposure¹⁴⁹ (/4 pts), language richness before the age of four as measured by diversity of language exchange contexts (/8 pts), current language exposure/use within the family (/16 pts), and current language use/exposure during different activities, e.g., reading, within an average week and in exchanges with playmates and family friends (/14 pts). The latter composite score also counts as an estimate of current language richness. An Exposure Index (/50 points) for L1 and L2 emerges by adding the above sub-scores. A visual representation of the relative contribution of each sub-score toward establishing the Exposure Index is given in **Figure 5**. As shown in **Figure 5**, current language use/exposure contributes the lion’s share (60%) to the calculation of the Exposure Index and, consequently, the Language Dominance Index. This converges with the findings of Bedore et al. (2012) in their large-

¹⁴⁶ Input and output are collapsed together (interaction from interlocutor to child and from child to interlocutor).

¹⁴⁷ Five additional points were allotted to number of years in elementary school as part of the exposure indices in France, but not in Germany, where children join elementary school between the age of 6 and 8.

¹⁴⁸ Total length of exposure (LoE) is calculated by subtracting age of onset of systematic sustained exposure to the respective language from the chronological age.

¹⁴⁹ Contrary to Unsworth (2015) and Unsworth et al. (2018), PaBiQ’s language use estimates were not only limited to the “inside home context” but also cover the “outside home context”.

scale study, where estimates of current language use (a composite score based on children's amount of exposure and language output) accounted for 60% of the variance in language dominance patterns of bilingual children.

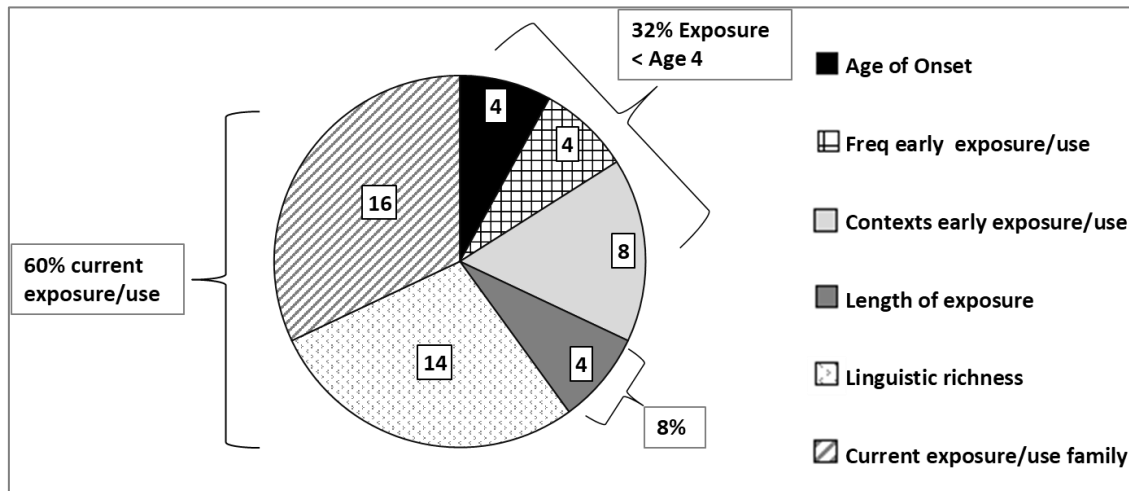


Figure 5: PaBiQ: Calculation of language exposure index

To calculate the LDI, the L1 Exposure Index is subtracted from that of the L2. This yields an estimate of L2 dominance on a scale from -50 (highly dominant in the L1) to +50 (highly dominant in the L2). De Almeida et al. (2017, p. 5) compared multiple LDI cutoffs around LDI = 0 (optimal balanced bilingual) against impressions of bilingual investigators of the individual children after interacting with them and their families in both of their languages and defined cutoff points for language dominance in an attempt to explore the use of this variable. An LDI between -5 and +5 was set as a cutoff separating dominant from balanced bilinguals. Children with LDIs ranging from -5 to +5 are classified as 'balanced'; children whose LDI is below -5 are considered dominant in the home language, while children with an LDI above +5 are classified as dominant in the societal language, German.

Socioeconomic Status

The questionnaire further allows for determining the family's socioeconomic status (SES) based on the mother's and the father's educational levels. Throughout this thesis, maternal rather than paternal educational level (as measured by years of education of the mother) is used as a metric for SES, as the former is reported to be a strong predictor of language development, especially for expressive vocabulary levels, in both

monolinguals (Hoff, 2003b; 2006) as well as child bilinguals (Calvo & Bialystok, 2014; Meir & Armon-Lotem, 2017; Paradis & Jia, 2017).

3.3 General procedure and statistical treatment

Testing took place in a quiet room of a kindergarten, a school, a speech-language therapy facility, or the child's home, with only the experimenter and the child present. It was split into three sessions á 40-55 minutes in the case of monolingual children and four sessions in the case of bilingual ones. All of the L1 tests were administered in a testing session exclusively dedicated to the L1 by linguistically trained research assistants who are native speakers of the respective languages. In the case of L1-dominant children, the non-verbal IQ test (the CPM) was also conducted in the child's L1 during the L1 session. The testing followed a predefined order that is outlined on the first page of the test project's protocol, see **Appendix IV**. Note that the protocol includes other tests that are not part of this thesis. At the end of each testing session, participants received a small reward for participating in the study (e.g., toys, stationery, stickers). The parental questionnaire was carried out with the parents/legal guardians of the participants through phone or face-to-face in their language of preference by trained native bilingual interviewers familiar with the respective culture. In addition to written protocols, children's oral responses were recorded using special recording devices, transcribed offline, verified, and scored by at least two linguistically trained research and student assistants.

Statistical analyses were conducted using R-Studio (R Core Team, 2012, 2016) and IBM SPSS 24, 26, and 27 (IBM Corp. released in 2016, 2019, and 2020, respectively). The method section of the respective chapters provides a detailed description of the specific statistical procedure employed in each of the studies presented in this thesis.

4 Study 1: Assessment of Heritage Language abilities in child bilinguals with and without SLI: Comparing a standardized L1 test battery to an Arabic LITMUS-Sentence repetition task¹⁵⁰

Abstract

To avoid misdiagnosis with specific language impairment (SLI) in bilingual children, it is recommended to evaluate both languages. However, unlike their monolingual peers, bilingual children acquire their heritage language under adverse input conditions. Focusing on Arabic, the study evaluates the clinical utility of a standardized test battery for Levantine Arabic and an Arabic LITMUS-sentence repetition task for ruling in/out SLI in bilingual children acquiring Arabic as a heritage language in Germany. Both tools were developed for the Lebanese bilingual context, in which Arabic is acquired in a diglossic context as a majority language. However, it is still unclear whether they can be reliably applied in heritage contexts.

Fifty-nine children (5;6-9;1) participated in the study: 18 bilingual typically developing (BiTD) children with early exposure to German (heritage BiTD), nine heritage bilingual children with SLI (heritage BiSLI), eight typically developing late successive bilingual children, and 24 typically developing monolingual Arabic-speaking children, who served as controls. Results showed that, unlike monolingual and late successive bilingual children with typical language development, many heritage BiTD children resembled their BiSLI peers, especially on measures of expressive/receptive vocabulary and morphosyntax production, which were negatively affected by early second language exposure. In contrast, the Arabic LITMUS-SRT did not disadvantage heritage BiTD children, who performed on par with their typically developing monolingual and late successive bilingual peers on most structures. BiSLI children, on

¹⁵⁰ This chapter contains substantial verbatim material from Abed Ibrahim (2025), originally published as an open-access article in *Journal of Speech, Language, and Hearing Research*; the article was published subsequent to the defense of this thesis and is reproduced here as part of the published dissertation. Full reference: Abed Ibrahim, Lina (2025). Assessment of Heritage Language Abilities in Bilingual Arabic-German Children with and without Developmental Language Disorder: Comparing a Standardized Test Battery for Spoken Arabic to an Arabic LITMUS Sentence Repetition Task. *Journal of Speech Language and Hearing Research*, 68(3S), 1569-1601. https://doi.org/10.1044/2024_jslhr-23-00617

the other hand, performed poorly on all structures. These results imply that sole reliance on standardized tests assessing language domains sensitive to reduced heritage language input bears a risk of overdiagnosis with SLI in heritage contexts. The Arabic LITMUS-SRT, on the other hand, provides a fair estimate of heritage language abilities. However, task construction should consider morphophonological aspects vulnerable in heritage Arabic.

4.1 Introduction¹⁵¹

To ensure a reliable diagnosis with SLI, it is recommended to assess a bilingual child in both languages since a genuine language disorder affects any language that is being acquired (ASHA, 2023; IALP, 2011). However, unlike their monolingual peers, bilingual children typically acquire their heritage language (HL) in a minority context under adverse input conditions with limited or no access to formal HL education (Montrul, 2008; 2016a). Consequently, they often experience a decline in HL abilities upon the shift to the dominant societal language during school years and develop no or very limited HL literacy skills (Montrul, 2016b). Moreover, depending on language typology, HL interlanguage features might overlap with clinical markers of SLI in monolingual contexts, limiting the diagnostic potential of the HL (Blom et al., 2019, 2022). Nevertheless, not all HL domains are equally affected by reduced input conditions, and some HL domains/properties appear to be more robust than others (Chondrogianni, 2023; Montrul, 2023). In general, structural properties of the language that are “rule- or constraint-based” are reported to be less sensitive to variations in HL input compared to language domains requiring a significant amount of input, such as vocabulary and functional morphology (Chondrogianni, 2023). For example, Scheidnes (2020) reported that bilingual children acquiring French in a minority immersion context performed better on *wh*-questions and relative clauses than on gender marking and verbal morphology in French. This is a diagnostically encouraging finding since syntactically complex constructions constitute a cross-linguistically valid vulnerable domain in SLI (Arosio et al., 2011; Friedmann et al., 2009; Hamann & Tuller, 2014). On the other hand, Correia et al. (2024) showed that bilingual children acquiring European Portuguese as an HL

¹⁵¹ Due to the cumulative nature of this dissertation, some overlap with previous sections is not avoidable.

encountered difficulties with object *wh*-questions indicating an interaction with language-specific typological features. This raises the question of whether and how HL assessment can contribute to ruling in/out the language impairment in heritage bilingual children and, crucially, what type of assessment tools should be employed for the assessment of HL abilities with the least possible overlap between heritage bilingual children with and without SLI.

One crucial step in this quest lies in finding appropriate tools for assessing bilingual children's abilities in their (weaker) HL that do not rely much on exposure or on literacy skills. Ideally, the test should not disadvantage bilingual children: it should assess multiple language domains, account for bilingual experience, and target language properties that represent an area of strength in typical bilingual acquisition but are vulnerable in children with SLI. However, bilingually norm-referenced tests are rarely available for heritage languages, and their clinical utility is largely limited to the specific bilingual contexts for which they were developed (e.g., Gagarina et al., 2010). By contrast, LITMUS sentence repetition tasks (see 2.3.1.1), which are available in more than 20 languages, were designed with bilingual acquisition in mind and have been shown to distinguish bilingual children with and without SLI not only in majority languages (e.g., Dutch: van Barreveld et al., 2023; French: Tuller et al., 2018; German: Hamann & Abed Ibrahim, 2017) but also in some heritage languages (e.g., Farsi: Komeili et al., 2020; Russian: Meir et al., 2016). Nevertheless, some studies indicate that, depending on language typology, certain constructions may be less suitable for assessing SLI in heritage bilinguals (e.g., Correia et al., 2023, for HL European Portuguese; Makrodimitris & Schulz, 2021, for HL Greek). Hence, the question arises whether standardized L1 tests can be used to identify SLI in the heritage language and how they compare with LITMUS sentence repetition tasks. To address this issue and inform bilingual assessment practice, this study examines the clinical potential of HL assessment in pre- and primary school-age bilingual children with Arabic as an HL and German as a societal language.

Arabic is of particular relevance for the German context since it constitutes the third most frequently spoken minority language in Germany, following Turkish and Russian (Federal Statistics Office/DESTATIS, 2023), with increasing numbers following the arrival of over 800000 asylum seekers of Syrian origin since 2015. However, finding appropriate diagnostic tools for Arabic as a heritage language is a challenging endeavor. First, Arabic is known for its diglossia and dialectal diversity (Furgeson, 1959; Watson,

2011), and HL tests are not available for all dialects. Second, when Arabic is acquired as an HL in minority contexts, acquisition takes place in a nondiglossic context (Albirini, 2016; Albirini & Benmamoun, 2022; Khamis-Dakwar & Khattab, 2014; Khamis-Dakwar & Makhoul, 2014; Saiegh-Haddad & Henkin-Roitfarb, 2014), where HL children are only exposed to the colloquial variety in the immediate home environment with limited or no access to the Standard variety. As a result, “they do not have access to the whole spectrum of functions” in which both varieties can be used and demonstrate very limited literacy skills in Arabic (Albirini & Benmamoun, 2022, p. 367). Hence, traditional assessment tools focusing on Modern Standard Arabic (MSA) and Arabic literacy are not suited for reliable assessment of Arabic HL abilities.

Due to the significant discrepancies among Arabic dialectal varieties (Watson, 2007) and the effect that this can have on diagnosis of developmental disabilities (Saiegh-Haddad, 2020; Saiegh-Haddad & Armon-Lotem, 2024), this study focuses on Levantine Arabic (Jordanian, Lebanese, Syrian, and Palestinian Arabic), which is spoken by a significant proportion of heritage Arabic speakers in Germany. In search of assessment tools for Levantine Arabic, the standardized test battery *Evaluation du Langage Oral chez l'enfant Libanais* (ELO-L; Zebib et al., 2017) and the LITMUS-SRT for Levantine Arabic (Zebib et al., 2020a) were considered as they showed promising results for disentangling SLI from typical language development in the Lebanese bilingual context (Zebib et al., 2019, 2020a). However, in Lebanon, Arabic enjoys a majority status and is acquired in a diglossic context. The question thus arises whether the ELO-L and the Arabic LITMUS-SRT can reliably distinguish typical from atypical bilingual language development in bilingual children acquiring Arabic in a heritage-minority context. To address this, the study compares the performance of three groups of Arabic–German bilingual children with typical and atypical language development on the ELO-L test battery and the Arabic LITMUS-SRT to evaluate whether either can reliably distinguish typical from atypical bilingual language development in children acquiring Arabic in a heritage bilingual context, in which Arabic may not (or no longer) be their dominant language.

In what follows, this chapter reviews relevant aspects of child heritage language development. Subsequently, it describes basic features of Arabic, reports on previous findings on the development of Arabic as a heritage language in nondiglossic minority contexts, and highlights aspects that can overlap with language markers of SLI in

monolingual Arabic-speaking children. Next, it presents the aims of the study and the methods, proceeds to describe the results, and ends with a discussion and conclusion.

4.2 Child heritage language development

Heritage child bilinguals are simultaneous or early successive bilinguals who acquire their HL via naturalistic input from birth alongside the ethnolinguistically dominant societal language, also called second language (L2; Benmamoun et al., 2013; Montrul, 2016a; Polinsky, 2006, 2018). In such a sociolinguistic context, the HL typically represents a sociopolitical minority language, and “proficiency in the [heritage language] is often valued less by dominant societal norms than [...] proficiency” in the societal language (Blom et al., 2019, p. 113; Montrul, 2016a). A crucial aspect of heritage childhood bilingualism is that exposure to the societal language starts in early childhood before the age of 5 years (Schwartz, 2004; Unsworth, 2005). According to Schmid and Köpke (2013), a stable state of HL grammar is reached when children are about 12 years old. This means that heritage children experience a significant reduction in the amount of HL exposure during a “crucial period of language development,” that is, at an age where “[t]he acquired linguistic competence is [still] not very stable” (Montrul, 2016b). In most cases, exposure to the HL is confined to the home environment with a limited number of interlocutors (Montrul, 2016a). Upon kindergarten entry, intense exposure to the societal language begins, and HL input is substantially reduced. In the absence or very limited formal instruction in the HL, the societal language dominates in frequency and domains of language use, progressively resulting in a dominance shift toward the societal language within 2–3 years of exposure to the societal language (Montrul, 2016a; Oller et al., 2011). With limited exposure and dominance shift, “the typical route of language development is interrupted in specific aspects of heritage language grammar (e.g., inflectional morphology),” but not in others, for example, phonology and core-syntax (Montrul, 2023, p. 401). As a result, heritage child bilinguals often develop asymmetric language profiles demonstrating monolingual-like performance in some aspects while showing signs of protracted development, differential acquisition, that is, incomplete acquisition or attrition in others (Montrul, 2008, 2016a; Schmid & Köpke, 2013). Both attrition and incomplete acquisition are argued to arise due to a shift in language dominance towards

the societal language (Montrul, 2008) and to be linked to cross-linguistic influence (Schmid & Köpke, 2017, but see also Flores, 2017). Manifestations of this include simplifications, reductions, reanalyses, and resorting to unmarked universal values (Montrul, 2008; see Benmamoun et al., 2013 for an overview). Recently, the use of the term incomplete acquisition has been subject to criticism by some researchers, who raised concerns regarding its theoretical appropriateness and the negative connotations associated with it (Domínguez et al., 2019; Kupisch & Rothman, 2018; Scontras et al., 2018). Specifically, the latter authors argue that the variability in HL development is an anticipated outcome of the variation bilingual children experience and should not be viewed as “incomplete” but rather a part of the “intra-individual variation within native grammars” (Scontras et al., 2018, p. 22). Benmamoun et al. (2013) and Montrul (2016a) suggested that it would be more appropriate to characterize HL grammars as divergent, where some aspects of the HL grammar are (partially) acquired without mastery under reduced input conditions (see Berman, 2004, for a discussion of the concepts “emergence”, “acquisition”, and “mastery”).

4.2.1 Sources of individual variation in heritage language development

Research on bilingual language acquisition has identified multiple mutually nonexclusive child-internal and language environment factors contributing to individual differences in HL development and maintenance (see Armon-Lotem & Meir, 2019; Kaltsa et al., 2015; Montrul, 2016a, 2016b; Paradis, 2023; Soto-Corominas et al., 2022 for overviews). The subsequent section elaborates on a selection of child internal and language environment factors that are of direct relevance to the current study.

4.2.1.1 Child-internal factors

Chronological age constitutes an important child-internal variable in language acquisition. In monolingual acquisition, language abilities grow as a function of chronological age. With increasing chronological age, the cumulative exposure to a language increases, and the cognitive abilities necessary to process structurally complex aspects of language mature. However, findings concerning the effects of chronological age on HL development are mixed. While some studies report an improvement in HL abilities with age (e.g., Armon-Lotem et al., 2011; Flores et al., 2017; Gagarina &

Klassert, 2018), others found no improvement or even negative effects of chronological age on HL performance (e.g., Armon-Lotem et al., 2021; Chondrogianni & Schwartz, 2020; Rodina & Westergaard, 2017). Investigating the differential contribution of chronological age to HL development is a very challenging endeavor and requires controlling for diverse language environment factors known to influence HL development. Indeed, many studies found that the effects of chronological age were counteracted by differences in the amount of HL input (e.g., Chondrogianni & Schwartz, 2020; Daskalaki et al., 2019; Rodina & Westergaard, 2017). Effects of age are also modulated by the nature of the respective linguistic structures and their timing in monolingual acquisition (Tsimpli, 2014). Hence, heritage bilinguals are expected to perform on par with their monolingual peers on early-acquired phenomena but not on late-acquired phenomena (Albirini, 2018; Makrodimitris & Schulz, 2021).

Age of onset (AoO) of bilingualism has been identified as an important child-internal variable in HL acquisition. Depending on the AoO of systematic exposure to the societal language/L2, child bilinguals can be divided into (a) simultaneous bilinguals if they start acquiring the societal language before the age of 3 years, (b) early successive (or sequential) child bilinguals if AoO lies between ages 3 and 4 years, and (c) late successive child bilinguals if the first exposure to the L2 takes place between ages 5 and 8 years (Meisel, 2009). The AoO of bilingualism is considered a crude measure of quantifying HL input since it coincides with the start of significant changes in the quantity and quality of HL input and is, in this sense, an “index of cumulative exposure to the HL” (Soto-Corominas et al., 2022, p. 4). Moreover, it serves as an index of the child's linguistic developmental stage and cognitive maturational level at the time she became bilingual (Armon-Lotem et al., 2021). Concerning AoO, there is a consensus in HL research that an older AoO to the L2 is beneficial for the development and maintenance of HL abilities (Kupisch, 2019; Kupisch & Rothman, 2018; Montrul, 2008, 2016a). An earlier exposure to the majority language entails divided input at an earlier age and thus less cumulative input in the HL. This might affect the acquisition of some linguistic properties that require more exposure to stabilize, especially those that are complex and involve opaque form–function mappings, or when their functional features are mapped differently in the HL and L2 or are absent in one of them, for example, nominal case and gender morphology (Thomas et al., 2014). The AoO of bilingualism also interacts with timing in monolingual acquisition, that is, the age at which a particular structure is acquired by monolingual

children (Tsimpli, 2014). According to the timing hypothesis, different linguistic structures have different maturational schedules. Some grammatical properties are acquired earlier than others in monolingual acquisition, that is, before the age of 4 years, for example, word order properties. In contrast, some aspects of complex syntax, semantics, pragmatics, and interface phenomena that reside outside “narrow syntax” are usually acquired later. For HL development, this means that if a given structural property is acquired late, that is, by age 6 or 7 years, the child might require more time to reach the “critical mass” of input needed to master that structure or never acquire it at all (Montrul, 2022). Hence, sequential bilinguals might have an advantage over simultaneous ones for late-acquired structures as they experience a longer period of monolingual exposure to their HL and have the opportunity to “stabilize at least some aspects of the HL” before intense exposure to the majority language begins (Kupisch, 2019, p. 7).

Cognitive abilities constitute a further child-internal factor that is involved in language learning. The latter includes, *inter alia*, analytic reasoning skills, working memory capacities, and executive functions (Paradis, 2023). Cognitive capacities are reported to be particularly implicated in (but not limited to) the acquisition of complex syntax (e.g., Zebib, Tuller, et al., 2020). Bilinguals constantly experience additional processing demands due to the concurrent activation of both the HL and L2 systems, which is overtaxed when they have to operate in their nondominant language based on reduced input and possibly weaker representations (Polinsky & Scontras, 2020). Consequently, depending on the ages of the participants, heritage bilinguals are likely to encounter difficulties with linguistic phenomena associated with higher processing costs, such as interface phenomena (e.g., pro-drop), low-frequency items, and structures involving long-distance dependencies such as relative clauses (e.g., Jia & Paradis, 2020).

4.2.1.2 Language environment factors

A substantial body of research has explored the differential effects of quantity and quality of language input on HL development. Quantity of exposure refers to the actual amount of language exposure a bilingual child receives in a given language. The most straightforward quantitative input factor is the length of exposure (LoE), which measures the approximate number of years since the onset of sustained exposure (Armon-Lotem & Meir, 2018). However, LoE does not capture day-to-day variations in the amount of exposure to each language over the years and does not consider how linguistic input is

distributed among various contexts and interlocutors. A widely used quantitative measure that accounts for variations in the amount of exposure in the two languages is the “relative amount of exposure.” It is determined by calculating the (average) amount of exposure a child receives in each of her languages inside and outside the home environment on a daily or a weekly basis and provides “an approximation of the exposure to each language relative to the other” in the form of percentages (Armon-Lotem & Meir, 2018, p. 195). The latter information is usually obtained via detailed parental questionnaires documenting language input/output in different contexts and with different interlocutors. Variations in input quantity have been shown to be an important predictor for the development and maintenance of HL abilities in HL morphosyntax, vocabulary, and even phonology (e.g., Chondrogianni & Schwartz, 2020; Correia & Flores, 2017; Daskalaki et al., 2019; Gollan et al., 2015; Hamann et al., 2020; Jia & Paradis, 2020). Crucially, a certain exposure threshold appears to be necessary to achieve functional bilingualism, although the threshold might vary for comprehension and production and differ by language domain (Gathercole & Thomas, 2009). Studies comparing HL to L2 acquisition have shown that the acquisition of the HL is much more dependent on its use at home than the majority L2, especially upon enrollment in kindergarten or school when systematic contact with the L2 begins (Paradis et al., 2021b; Sorenson Duncan & Paradis, 2020).

Many studies documented positive associations between the quality of the HL input and proficiency in the HL (e.g., Albirini, 2014; Jia & Paradis, 2020). Quality concerns the richness and the nature of the HL input bilingual children receive (e.g., the variety of language, diversity of input sources), which in many cases substantially differ from the input received by monolingual children acquiring their native language in a majority context (Montrul, 2016a). The richness of the HL environment can vary as a function of diversity and frequency of HL-promoting activities, such as literacy-related activities, audiovisual media, extracurricular activities, and socializing with playmates in the HL (Paradis, 2023). An important factor affecting the quality of HL input is the parental generation. Parents of heritage bilingual children serve as the primary source of HL input in heritage minority contexts and might themselves have low HL proficiency, lack HL literacy skills, or be speakers of (contact-induced) varieties that have undergone dialectal leveling, as in the case of Immigrant Turkish in Germany (Chilla, 2022; Chilla & San, 2017; Daskalaki et al., 2023, 2024; Rehbein et al. 2009). Even first-generation

immigrants, who tend to have stronger HL abilities relative to second-generation immigrants, might display varying degrees of HL proficiency and language attrition, especially in interface structures (Köpke & Genevskaja-Hanke, 2018). For example, Otheguy and Zentella (2012) and Albirini et al. (2011) reported an overuse of overt pronouns in pragmatic contexts that favor the pro-drop strategy by speakers of Spanish and Arabic as heritage languages, respectively. Daskalaki et al. (2019, 2020) studied subject placement in wide focus versus embedded interrogatives in HL Greek in Western Canada, comparing school-age monolingual Greek-speaking children and their parents with Greek-English bilingual children and their parents (first- and second-generation migrants). They found that bilingual parents and their children produced significantly more preverbal subjects in contexts requiring postverbal subjects than the monolingual groups. Importantly, this tendency appeared to be more pronounced in second-generation families than in first-generation families.

The quality and also the quantity of HL input are also influenced by broad distal language environment factors such as the parents' socioeconomic status (SES), which serves as an index of “overall family environment and cultural capital” (Paradis et al., 2021). SES is frequently operationalized by the levels of parental education and occupation. Various studies report a connection between levels of maternal education and language performance of bilingual children in the societal language (e.g., Meir & Armon-Lotem, 2017). However, the relationship between levels of maternal education and bilingual children's HL abilities appears to be modulated by other factors, including the language of parental education (Paradis et al., 2021; Sorenson-Duncan & Paradis, 2020).

A further broad distal factor is the sociopolitical status of the HL as a majority versus minority language, which affects the availability of schooling opportunities in the HL. A significant qualitative difference between the development of the HL in minority versus majority contexts arises due to the heritage speakers' limited access to formal education in the HL, especially during the crucial school-age period (Kupisch & Rothman, 2018; Montrul, 2016a). The lack of access to formal instruction and literacy-promoting activities in the HL affects the quantity and, crucially, the quality of HL input. Literacy in the HL is assumed to contribute to HL maintenance at different levels. At a psycholinguistic and neurolinguistic level, literacy is argued to influence the cognitive organization of language in the brain by supplementing existing lexical and phonological representations with orthographic representations, thereby adding new synaptic

connections (Köpke, 2004; Köpke, 2007). From a sociolinguistic perspective, HL literacy can have a positive impact on HL maintenance by enabling HS speakers to remain connected to the HL via reading, which is likely to serve as a crucial “source of confirming evidence for the L1/[HL]” (Köpke, p. 11; see also Andreou et al., 2020, and Sharwood Smith & van Buren, 1991). Supporting evidence can be found in studies showing that HL literacy promotes long-term maintenance of the HL (e.g., Albirini, 2014; Bayram et al., 2019; Cummins, 2005; Eisenclas et al., 2013; Jia, 2008; Zaretsky & Bar-Shalom, 2010). A further effect of literacy is that it can serve as a countervailing factor to language change in heritage contexts (Rothman, 2009). The limited “contact with formal registers is intrinsically paired with exposure to input rich in linguistic variation” (Flores, 2015, p. 254), which is a natural property of language (Labov, 1972). Hence, HL grammars tend to be less affected by countervailing standardization processes, and the linguistic variation, typically found in colloquial language registers of the monolingual community, is likely to be “preserved or even [accelerated and quantitatively] amplified” in the language of heritage speakers and their children (Flores & Rinke, 2020, p. 26; see also Flores, 2015; Silva-Corvalán, 1986).

4.2.2 HL outcomes across linguistic domains

Although heritage bilinguals are considered native speakers by definition (Rothman & Treffers-Daller, 2014), compared to their nonheritage monolingual age-matched peers, they often present uneven language profiles and rarely display age-appropriate proficiency levels in all linguistic domains of their HL, especially when the HL represents their weaker language (Albirini & Benmamoun, 2014; Benmamoun et al., 2013; Montrul, 2008). Phonology represents a relative area of strength in HL acquisition. Heritage bilinguals are frequently reported to have a rather “authentic” pronunciation and have been shown to demonstrate more advanced phonetic/phonological abilities relative to late and adult L2 learners (Kupisch et al., 2021; Saadah, 2011). This robustness is argued to reflect the advantage of early exposure to the HL, as many phonetic and phonological properties are acquired early in monolingual acquisition and should be less susceptible to effects of reduced input conditions (Polinsky & Scontras, 2020). However, this advantage pertains to the segmental aspects of HL phonology (e.g., the ability to contrast and categorize sounds), and subtle differences between the heritage and the monolingual

baseline variety may exist in the phonetic values of vowels and consonants, for example, voice onset time and vowel quality (Kupisch et al., 2021; Saadah, 2011). Suprasegmental aspects, on the other hand, are reported to be particularly vulnerable in HL acquisition (e.g., Kim, 2020) and more prone to cross-linguistic influence than segmental phonology (Lleó, 2018).

With respect to HL lexical abilities, ample research has shown that child heritage bilinguals lag behind their monolingual age peers on HL receptive and expressive vocabulary measures, especially when vocabulary breadth (number of words in the child's lexicon) is assessed (Bialystok et al., 2010; Czapka et al., 2021; Montanari et al., 2019). This vulnerability has been ascribed to the distributed characteristic of vocabulary knowledge across the domains of two languages (e.g., home vs. school) and its heavy reliance on piecemeal learning, making it particularly sensitive to restricted input effects (Montrul & Foote, 2014). Lexical items are “learned independently in each language regardless of whether or not the relevant concept has been acquired” (Unsworth, 2016, p. 109). Hence, unless some overlap exists between the HL and the L2 vocabularies in the form of cognates, lexical bootstrapping between the two languages remains limited. A commonly observed pattern is age-appropriate vocabulary development during preschool years, followed by stagnation as a result of extensive exposure to the majority language via schooling (Montrul, 2016a). Moreover, many studies report a discrepancy between production and comprehension for HL vocabulary. For example, in a sample of 49 French–English simultaneous bilinguals ($M_{\text{age}} = 5;0$ [years;months]), Thordardottir (2011) found expressive vocabulary to be more susceptible to effects of reduced exposure than receptive vocabulary: While balanced bilinguals performed on par with monolingual children on receptive vocabulary measures in their first language, French, only those children with exposure levels to French $> 60\%$ demonstrated comparable performance to their monolingual peers on expressive vocabulary measures. Contrary to vocabulary's breadth, vocabulary's depth (quality of lexical entries), for example, expressive word classes, has been reported to be more robust to effects of reduced input as it draws upon conceptual representations of the lemmas that are shared across the languages and involves metalinguistic knowledge that represents a shared skill between the languages (Armon-Lotem et al., 2021).

Functional morphology represents another locus of difficulty for child and adult HL speakers from a range of typologically different languages (Montrul, 2018). In

particular, morphological aspects involving non-transparent form-function mappings appear to be susceptible to reduced input, for example, number and gender agreement on nouns in Spanish (Scontras et al., 2018), plural morphology of Arabic and Welsh (Albirini, 2018; Saiegh-Haddad et al., 2012; Thomas et al., 2014), the gender systems of Russian and Dutch (Rodina & Westergaard, 2017), and case morphology in Greek and Russian (Chondrogianni & Schwartz, 2020; Meir & Armon-Lotem, 2015). Common error patterns include producing syncretic forms (i.e., ambiguous morphological forms such as case marking in German), default forms, omissions of inflectional markings, or simplifications by reducing the number of distinctions of the inflectional markings (Polinsky & Scontras, 2020). The latter two error patterns are also attested in monolingual children with SLI acquiring inflectionally rich languages (Leonard, 2014b).

Although (morpho)syntax is also influenced by fluctuations in HL input, available research suggests that it is more resilient to reductions in HL input compared to vocabulary and functional morphology (Benmamoun et al., 2013; see Lohndal et al., 2019, for a discussion). In particular, early-acquired phenomena involving macroparameters (core syntactic structures), such as basic word order and head directionality, are argued to require less input to stabilize (Albirini et al., 2011; Kaltsa et al., 2015). However, syntactically very complex structures and structures that stabilize late or very late in monolingual acquisition in a given language are reported to be vulnerable in HL acquisition. Such structures involve language-external cognitive resources such as working memory and are associated with “increased semantic complexity and/or discourse properties that are more demanding than narrow syntactic properties,” that is, interface phenomena. Hence, they require a significant amount of input to stabilize (Tsimpili, 2014). In heritage contexts, late-acquired properties are acquired at an age at which heritage bilingual children typically experience drastic reductions in HL input due to immersion in mainstream education in the majority language, which, in many cases, has become their dominant language. In particular, structures at the syntax–pragmatics (e.g., null and overt pronouns) and some aspects at the syntax–semantics interface (e.g., binding interpretation of anaphors in long-distance configurations) are reported to be more sensitive to variations in HL input than syntactically conditioned ones (Argyri & Sorace, 2007; Daskalaki et al., 2020) and tend to show protracted development and cross-linguistic influence from the dominant language. For example, in child heritage Bulgarian, Genevskaya-Hanke and Hamann

(2024) reported an overuse of lexical determiner phrases and low use of pronominal subjects in German-dominant heritage Bulgarian children, which they attributed to cross-linguistic influence from the dominant language German, in which topics can be positionally marked. A marked difficulty has also been reported for structures involving long-distance nonlocal dependencies, especially for noncanonical complex syntactic constructions (e.g., object relative clauses in German) whose disambiguation heavily relies on morphological cues (case, number, and gender), which represent a vulnerable domain in HL acquisition (Polinsky & Scontras, 2020). For example, Chondrogianni and Schwartz (2020) showed that Greek heritage bilingual children of primary school age encountered difficulties in integrating case marking cues for the interpretation of noncanonical object–verb–subject (OVS) structures and tended to rely on linear word order for thematic role assignment.

Taken together, language domains requiring a significant amount of input, such as vocabulary, nominal and verbal morphology (Montrul, 2016a, 2016b), and late-acquired linguistic phenomena (Flores et al., 2017; Tsimpli, 2014), appear to be particularly affected by reduced input conditions and are likely to lead to diagnostic confounds in heritage contexts. In contrast, certain aspects of morphosyntax appear to be more resilient to the effects of reduced input conditions and might thus be exploited for diagnostic purposes in heritage contexts.

4.3 Arabic as a heritage language

4.3.1 Some notes on Arabic

Arabic is a Semitic root-derived¹⁵² synthetic language with rich non-concatenative¹⁵³, that is, non-linear, morphology that is based on a system of roots and templatic patterns (Ryding, 2005; Saiegh-Haddad & Henkin-Roitfarb, 2014). An important sociolinguistic feature of Arabic is diglossia (Ferguson, 1959), where two related varieties of the same language are systematically employed within the same speech community depending on the communicative function, that is, the sociolinguistic context. One can distinguish between two main varieties of Arabic: MSA, which is based on Classical Arabic (“*fuṣḥā*”) and is used in religious ceremonies, conventional reading and writing, media and in formal speech, and colloquial or vernacular Arabic (“*ammiya*”), which is a broad term covering “a wide variety of regional and local, mainly spoken, dialects across the Arabic-speaking region” (Albirini & Benmamoun, 2022, p. 362). Colloquial Arabic is predominantly implemented in everyday oral communication at home and within the community (Albirini & Benmamoun, 2022; Bassiouney, 2009).

Colloquial Arabic is characterized by its dialectal diversity with more than 27 dialects, all of which are structurally related to MSA (Alsudais et al., 2022; Watson, 2011), a rather uniform variety, which serves as a standard mutually intelligible language for the literate Arabic-speaking community being the only language of literacy across the

¹⁵² Within Arabic root-pattern morphology, the root, which typically consists of an invariable triliteral consonantal sequence (some roots comprise 4 to 5 consonants), provides the core semantic meaning of all lexemes belonging to the same root-related family. The pattern (verbal or nominal), on the other hand, is made up of a vocalic skeleton and affixation consonants (Saiegh-Haddad & Henkin-Roitfarb, 2014), and conveys “the word’s prosodic structure as well as its syntactic category and related grammatical properties” (Vaknin-Nusbaum & Saiegh-Haddad, 2020, p. 2418; see also Haddad et al., 2018). The root’s consonantal constituents are called “radicals” (C₁, C₂ and C₃) and are discontinuously inserted into designated slots within the templatic (fixed prosodic) pattern in order to form new words and stems. This is exemplified by the Arabic triconsonantal root *KTb*, which is associated with the concept of “writing”. Applying root-pattern morphology to it results in various semantically-related words, e.g., *kataba* ‘he wrote/ has written’, *maktaba* ‘library/ bookstore’, *kita:b* ‘book’, *ka:tib* ‘author/ writer’ (the letters in bold represent the consonantal root). In order to look up an Arabic word in a dictionary, it is necessary to know its semantic root, as all the lexemes related to it are listed under the root. Except for particles and non-Arabic roots, “the citation from any root is the 3rd sing masc. [perfective] verb” (Badawi et al., 2004: p. 26).

¹⁵³ Note that Arabic inflectional morphology also has concatenative forms that involve suffixation, e.g., regular masculine and feminine plural forms (i.e., sound plurals).

Arabic-speaking regions, and placing linguistic distance between the vernacular and the standard at the heart of literacy learning (Saiegh-Haddad, 2018, 2022). Despite being linguistically related to the standard variety, spoken colloquial varieties differ from MSA at the phonological, morphological, morphosyntactic, and lexical–semantic levels, where the linguistic distance is most pronounced in the domains of phonology and lexicon (Saiegh-Haddad, 2004). The extent to which the local varieties diverge from the Standard variety varies from country to country. For this reason, the focus of this study is restricted to Levantine Arabic varieties (Syrian, Lebanese, Palestinian, and Jordanian) as they bear a considerable resemblance at all linguistic levels (Aoun et al., 2010). At the structural level, Levantine Arabic varieties predominantly employ SVO (default word order) and VSO word orders and have an elaborate subject–verb agreement paradigm in terms of person, number, and gender, though less elaborate than MSA. Similar to MSA, Levantine Arabic is a null-subject language allowing the omission of the subject in tensed clauses (pro-drop strategy). Levantine Arabic uses both concatenative (e.g., dual and regular plural forms) and nonconcatenative (e.g., broken plurals) morphological forms; however, the systems are less elaborate than MSA or might use different morphemes. Moreover, unlike MSA, Levantine Arabic lacks case markings on nouns and adjectives but has a richer aspectual system with independent morphological markers, for example, the proclitic “*ʕam*,” which marks the progressive aspect in the present tense (Ryding, 2005). Despite subtle differences, Levantine Arabic resembles Standard Arabic in the use of definite articles as well as in subcategorization and relativization rules (Aoun et al., 2010).

4.3.2 Previous findings on the acquisition of Arabic as a heritage language

A crucial aspect of Arabic HL acquisition is that, unlike typical monolingual Arabic language acquisition (Saiegh-Haddad, 2003), acquisition does not take place within a diglossic context (Khamis-Dakwar & Makhoul, 2014; Saiegh-Haddad & Henkin-Roitfarb, 2014). In monolingual contexts, Arabic-speaking children experience how Standard and colloquial Arabic varieties “co-exist and function in the Arabic linguistic landscape” (Albirini & Benmamoun, 2022, p. 366). Children naturally acquire the local colloquial variety at home and are exposed to the Standard variety through formal instruction at school. In contrast, heritage children are predominantly exposed to the spoken variety in a rather restricted familial context, while exposure to Standard Arabic,

which represents a third language for them, is either minimal or nonexistent (Albirini, 2018). This contextual restraint has a stark impact on their “knowledge and use of various aspects” of colloquial and Standard Arabic and “their awareness of how diglossia transpires in everyday interactions and communication” (Albirini & Benmamoun, 2022, p. 366).

Previous studies on heritage Arabic showed that adult heritage speakers show gaps relative to monolingual native speakers with respect to HL lexicon and verbal and nominal inflectional morphology, particularly gender and number agreement as encoded in subject–verb agreement and noun–adjective agreement (Albirini, 2014; Albirini et al., 2013). As a simplification strategy, they frequently opt for the default singular masculine form. They also struggle with dual noun forms and with nonconcatenative plural morphology, especially with irregular plurals with geminated and defective roots, for example, *raff* → *rfuuf* “shelves” (Albirini & Benmamoun, 2014; Benmamoun et al., 2014). As compensatory strategies, heritage speakers resort to default strategies attested in monolingual child language acquisition (e.g., Albirini, 2015), such as the use of less marked forms, such as the feminine sound plural (Saiegh-Haddad et al., 2012). With respect to sentence structure, Arabic heritage speakers with English as L2 are reported to preserve basic sentence structure of Arabic; however, they show a strong preference for the default SVO word order and overuse overt pronouns in contexts that pragmatically require using the pro-drop strategy (Albirini et al., 2011). The authors attribute this to cross-linguistic influence from English, which overtly marks the sentential subject and has invariable SVO word order. As to relativization, as is the case for many languages, a subject–object asymmetry is reported for heritage Arabic (Albirini & Benmamoun, 2014). Although both subject and object relatives involve movement and embedding, object relatives are considered computationally more complex (hence more taxing to working memory resources) than subject relatives because they feature a longer distance between the moved constituent and its extraction site, which is occupied by a resumptive pronominal clitic. In addition to that, the long-distance dependency in Arabic object relative clauses “requires [considering] the gender and number of the relativized lexical head and how they are morphologically represented on the resumptive pronoun” (Albirini, 2018; see also Aoun et al., 2010). Albirini and Benmamoun et al. (2014) found that heritage Arabic speakers encountered problems with gender and number feature matching on the resumptive pronoun and its antecedent in long-distance dependencies in

object relative clauses, such as restrictive relative clauses. With respect to phonology, heritage Arabic speakers' phonological abilities appear relatively spared, at least for core segmental aspects (e.g., phoneme inventories and basic segmental contrasts), despite subtle differences from monolinguals, particularly in the production and perception of phonetically marked contrasts such as emphatic vs. plain consonants (Al Omary, 2024; Saadah, 2011). However, individual outcomes can vary depending on factors known to influence heritage language acquisition.

Studies on Arabic HL development in children showed similarities to adult heritage speakers. For example, Albirini (2015) reported that 5- to 6-year-old heritage child bilinguals demonstrated lexical retrieval difficulties relative to their monolingual age-matched peers. Haddad (2022) reported superior expressive vocabulary abilities for 7-year-old bilingual children acquiring Arabic in Lebanon (majority context) compared to bilingual children acquiring Arabic in Sweden (minority context). With respect to inflectional morphology, Albirini (2018) showed that heritage bilingual children resembled heritage adults by frequently replacing verbal agreement markers with the singular masculine form, which represents an “underspecified form of verbal agreement morphology ... since [unlike singular feminine, dual, or plural forms] it does not require a specific affix for gender or number” (p. 193). Regarding plural morphology, the children in Albirini's (2018) study showed similar patterns to adults by replacing irregular nonconcatenative plural forms with the default most frequent semantically unmarked form, namely, the sound (regular) feminine plural. As to complex constructions, a subject–object asymmetry was observed for relative clauses, where heritage child bilinguals performed on par with their monolingual peers on subject relatives but showed poorer performance on the structurally more complex nonsubject restrictive relative clauses (direct object, indirect object, and oblique object relative clauses). As to predictors of performance, Albirini (2018) found an older age of exposure to the societal language to be positively associated with better performance on the latter morphosyntactic phenomena. With respect to word order, similar to heritage adults, heritage bilingual children also show a preference for default SVO word order. For example, Bos (1997) compared monolingual and bilingual Dutch–Moroccan children's comprehension of Arabic relative clauses with an alternating word order in the main clause (SVO vs. OSV). While bilingual children performed well on the basic SVO word order, they were outperformed by their monolingual peers on OVS structures. Her study also showed that

monolingual children were more efficient in processing morphological cues. Concerning phonology, although heritage children have been shown to generally converge with monolingual age controls at the segmental level, subtle differences to monolingual children can also be found, for example, for standard Arabic phonemes, voice onset time, and phonologically conditioned allomorphy, such as assimilation in the case of geminate consonants (Khattab, 2002).

The latter findings are relevant for assessing Arabic in heritage contexts, especially since some Arabic HL interlanguage features show overlap with clinical markers for SLI in monolingual contexts (Blom et al., 2019) and could lead to diagnostic problems. For example, monolingual Arabic-speaking children with SLI have difficulties using verb forms marked for tense and agreement and tend to use less marked (i.e., underspecified) verb forms, such as singular masculine verb forms instead of more marked feminine or plural forms, imperfective bare forms that resemble imperative forms (as a default tense form) in place of fully inflected (tensed) verb forms, and first-person verb forms instead of third-person verb forms (Abdalla & Crago, 2008; Aljenaie, 2010; Fahim, 2005, 2017; Taha et al., 2021a). With respect to nominal inflectional morphology, Abdalla et al. (2013) found that Kuwaiti Arabic-speaking children with SLI scored significantly lower on nonconcatenative plural morphology than their typically developing age-matched peers. However, although heritage typically developing bilingual children and monolingual children with SLI have problems with irregular plural morphology in Arabic, they show distinct error patterns. While heritage BiTD employ the default feminine plural suffix (concatenative plural form; Albirini, 2018), children with SLI mainly produce uninflected singular forms (Abdalla et al., 2013). With respect to complex morphosyntax, Arabic-speaking children with SLI are reported to encounter exceptional difficulties with the comprehension and production of noncanonical sentences featuring nonlocal dependencies, such as sentences with clitic left dislocation (Rakhlin et al., 2020; Shaalan, 2010; Taha et al., 2021b). Just as reported for many languages, Arabic-speaking children with SLI are shown to perform poorly on sentence repetition, especially on constructions featuring complexity aspects (e.g., embedding, *wh*-movement) known to be cross-linguistically vulnerable for populations with SLI. For example, Shaalan (2010) found subject and object relative clauses to be challenging for Arabic-speaking children with SLI. Taha et al. (2021b) administered a LITMUS-SRT for Palestinian Arabic to monolingual children with and without SLI (ages 4;10–6;10).

Although the SLI group showed lower accuracy scores on the task as a whole, including simpler constructions, a much larger performance gap was observed for sentences with higher levels of structural complexity. Interestingly, a distinct error pattern was observed for children with SLI, namely, the omission of multiple sentential elements resulting in a fragmented structure. The findings for sentence repetition imply that LITMUS-SRTs can be particularly useful for differentiating between typical and atypical development of Arabic as an HL, especially since most of the complex sentence types featured in LITMUS-SRTs, including the Arabic LITMUS-SRT investigated in this study, are not vulnerable in heritage Arabic. Testing the role of diglossia in assessment and diagnosis, Saiegh-Haddad and Ghawi-Dakwar (2017) reported greater difficulties in non-word repetition among Palestinian Arabic-speaking children with SLI than children with typical language development, with more prominent and developmentally persistent difficulties in children with SLI when nonwords encoding unique standard phonemes are targeted. Ghawi-Dakwar and Saiegh-Haddad (2025) also report greater word learning difficulties among children with SLI than typically developing children and larger difficulties in children with SLI for novel words with Standard Arabic phonology, at least in comprehension probes. This implies that assessment of SLI in HL contexts is best when it is based on the spoken variety, rather than on standard structures, to avoid over/under diagnosis, especially that heritage child bilinguals acquire Arabic in a nondiglossic context and are less likely to be familiar with MSA (Saiegh-Haddad, 2020; Saiegh-Haddad & Armon-Lotem, 2024).

4.4 The present study

The present study informs bilingual assessment practice by examining the clinical potential of HL assessment for accurate diagnosis of SLI in pre- and primary school-age bilingual children with Arabic as HL and German as societal language. Traditional assessment tools focusing on MSA and Arabic literacy are not suited for reliable assessment of HL abilities. Thus, it is necessary to assess children in the spoken dialect, which is used as an HL at home. This study focuses on Levantine Arabic, a hypernym referring to the Lebanese, Syrian, Jordanian, and Palestinian varieties. The latter dialects bear a considerable resemblance at all linguistic levels since they all belong to the Eastern

Arabic dialect group (Aoun et al., 2010). In search for standardized assessment tools for Levantine Arabic, the ELO-L standardized test battery (Zebib et al., 2019) was considered. ELO-L is the only comprehensive test battery for assessing oral Arabic language abilities that offers bilingual norms. Moreover, since LITMUS-SRTs, which feature complex syntactic constructions, have shown promising results for disentangling SLI from typical language development in bilingual children in the majority language, the study also evaluates the clinical utility of a LITMUS-SRT for Levantine Arabic (Zebib et al., 2020a).

ELO-L is one of the few available tests for assessing oral Arabic language abilities in bilingual children. It consists of five tests assessing receptive vocabulary, expressive vocabulary, sentence comprehension, sentence production, and phonology and showed excellent diagnostic accuracy for bilingual children acquiring Arabic in Lebanon (Zebib et al., 2019). However, it was normed based on bilingual populations acquiring Arabic in the Lebanese majority context and does not consider the bilingual experience in heritage contexts, where Arabic represents a minority language and is acquired in a nondiglossic context. Hence, the question arises whether ELO-L is suitable for the assessment of heritage bilingual children, especially as it includes subtests assessing vocabulary and functional morphology, which are reported to be sensitive to variations in HL input. Similar to the ELO-L, the Arabic LITMUS-SRT was developed to be employed in the Lebanese bilingual setting, that is, in a majority language context. However, it was constructed in accordance with the LITMUS principles (Marinis & Armon-Lotem, 2015) and includes structures with complexity aspects (embedding and *wh*-movement) that are vulnerable in SLI but should be robust in bilingual language acquisition. Hence, it is expected not to disadvantage bilingual children. Previous studies piloting the Lebanese version of the Arabic LITMUS-SRT showed promising results for bilingual children acquiring Arabic in the Lebanese majority context. Saad and Henry (2020) reported good diagnostic accuracy for the task using three different scoring measures: (a) identical repetition, that is, verbatim; (b) target structure, which assesses the preservation of the targeted construction disregarding lexical and morphological errors not affecting the realization of the targeted construction; and (c) grammaticality, which requires the response to be grammatical regardless of whether the targeted construction was produced or not. The latter study further found that the structurally more complex constructions, namely, finite complement clauses, subject relatives, and object relatives, better

discriminated BiTD from bilingual children with SLI (BiSLI) than the structurally simpler sentences. Interestingly, the scoring method “identical repetition,” which represents the simplest and fastest scoring procedure for speech-language pathologists, proved to be the most effective one. See Supplemental Material S1, Table 2, for examples of the test stimuli and of the scoring procedures. However, the latter study featured younger child bilinguals aged 5–6 years and was conducted with bilinguals acquiring Arabic in a majority context, leaving the question of whether it can be reliably employed to disentangle BiTD and BiSLI in heritage contexts and in older bilingual children.

The only study comparing the ELO-L and the Arabic LITMUS-SRT in a heritage bilingual context is Hamann et al. (2020), which is an extension of Abed Ibrahim et al.'s (2020) study. Hamann et al. (2020) compared the performance of older (ages 6;0–13;0) heritage typically developing bilingual Arabic–German children on the ELO-L and SRT to a group of age-matched, late successive, typically developing bilinguals within their first 3 years of exposure to L2-German. While the late successive bilingual group demonstrated good performance on all of the ELO-L's subtests, the results of the heritage bilinguals were rather discouraging. Heritage bilinguals showed comparable performance to late successive bilinguals only on subtests assessing morphosyntax comprehension and phonology, whereas a significant proportion of them performed in the pathology range on subtests assessing expressive/receptive lexical skills and morphosyntax production even after adjusting the cutoff points according to their language dominance. It should be noted that most of the participants of the heritage group in Hamann et al.'s (2020) study were German-dominant and had been exposed to German for more than 6 years. It hence remains to be answered whether younger heritage bilinguals with a shorter LoE to the L2 would also perform in the pathology range. With respect to the SRT, heritage bilinguals performed on par with late successive bilinguals. However, the task appeared to be rather easy for bilingual children older than 9 years, who showed ceiling performance on the task. Hence, it is still unclear whether the Arabic LITMUS-SRT can differentiate heritage bilinguals with typical language development from heritage bilinguals with SLI in pre- and primary school-age bilingual children.

4.4.1 Research questions and aims of the present study

The present study builds upon Hamann et al.'s (2020) study and expands it by investigating a larger sample of younger heritage BiTD and late successive bilingual children who are in preschool or primary school age, where language evaluation typically takes place in Germany. Importantly, the study includes a sample of bilingual Arabic–German children with SLI and a control group of typically developing, monolingual, Arabic-speaking children. The study addresses the following broad research questions. It first asks whether a standardized test battery and a LITMUS-SRT for Levantine Arabic can be reliably used for ruling in/out language disorder in heritage contexts. Second, the study aims to investigate the extent to which child-internal and language environment factors influence the performance of typically developing heritage bilingual children across the different linguistic domains assessed by the tests to determine which language measures are relatively robust against reduced input effects in typically developing children and can be exploited for diagnostic purposes. Finally, given that some studies showed that certain structure types are problematic for heritage speakers, for example, object relative clauses, the study further asked whether all constructions of the Arabic LITMUS-SRT are suitable for assessment of HL morphosyntactic abilities in heritage bilingual minority contexts. To answer these overall questions, the study asks the following more specific questions:

- i. How do bilingual heritage children with and without SLI compare to their monolingual and late successive typically developing peers in their Arabic abilities as estimated by the ELO-L and the LITMUS-SRT for Levantine Arabic? Specifically, can both tests be applied with equal success to rule in/out language impairment in heritage contexts?
- ii. Do child internal and language environment factors known to influence performance on language measures, such as chronological age, age of bilingualism, LoE to the L2, HL/L2 input quantity and quality, and SES, explain variance in performance of the bilingual typically developing children on the different subtests of the ELO-L assessing receptive/expressive vocabulary, morphosyntax comprehension/production and phonology (repetition of colloquial

and standard Arabic words) and the Arabic LITMUS-SRT? For the Arabic LITMUS-SRT, the study also asks whether performance is rather predicted by working memory measures or by language abilities as estimated by the ELO-L.

- iii. How do heritage BiTD compare to monolingual typically developing children (MoTD), late successive BiTD, and heritage BiSLI on the individual structures of the Arabic LITMUS-SRT, and which structures can reliably distinguish between BiTD and BiSLI in heritage contexts?

4.5 Method

4.5.1 Participants

A total of 59 children aged between 5;6 and 9;1 were included in this study: 26 BiTD, nine BiSLI, and a control group of 24 MoTD. All children were exposed to Levantine Arabic varieties from birth. Bilingual children, whose systematic exposure to L2-German began before the age of 5;0, were classified as heritage bilinguals (Montrul, 2016a; Schwartz, 2004). Children who were born in an Arabic-speaking country and migrated to Germany after the age of 5;0, that is, at school age, were not considered heritage bilinguals and were classified as late successive bilinguals and constitute a further control group. Given their relatively high AoO (> 5;0) and short LoE (< 3;0) to the L2, the latter children were expected to demonstrate strong HL maintenance, especially as they spent an extended period of time in a monolingual qualitatively rich language environment and probably had access to formal language registers (Daskalaki et al., 2023; Paradis et al., 2021b).

The monolingual typically developing group was recruited from three kindergartens and a primary school in Palestine, Lebanon, and Jordan. The bilingual data were collected within the projects “BiLaD” and “BiliSAT” in different parts of Germany and supplemented by the author’s own data. Bilingual typically developing children were recruited from kindergartens, primary schools, and community-based schools offering instruction in MSA, for example, weekend language schools. Children likely to have SLI, that is, with a clinical diagnosis of SLI, were recruited from specialized speech-language

therapy facilities and kindergartens with special inclusion programs. As inclusion criteria for bilingual participants, children had to have at least 18 months of exposure to the L2 (German) and demonstrate functional bilingualism.

The clinical status of the recruited bilingual children was verified via a comprehensive procedure using standardized assessment in the HL and L2, applying dominance-adjusted cutoff points as outlined in **3.1.1**. A bilingual child was assigned to the BiSLI group if she scored below language dominance-adjusted cutoffs in two language domains in each language (phonology, expressive/receptive vocabulary, and morphosyntax production and comprehension). Since expressive vocabulary is a notorious locus of difficulty for bilingual children, only receptive vocabulary was considered. Concerning monolinguals, only children who scored above the monolingual cutoff (-1.25 SD) in four out of the five ELO-L subtests were regarded as typically developing. A detailed participant overview is provided in **Table 5**. In this table, the bilingual typically developing children are categorized based on AoO to the L2 into heritage typically developing bilinguals (heritage BiTD, $n = 18$) and late successive typically developing bilinguals (IL2 BiTD, $n = 8$). The heritage BiTD were either born in Germany (13 out of 18) or immigrated to Germany with their parents at an early age (five¹⁵⁴ out of 18). Their AoO to German varied from ages 0;0 to 4;2, resulting in simultaneous ($n = 5$) or early successive ($n = 13$) child bilingualism. Four of the heritage BiTD attended Sunday schools, but only two of them had literacy skills in Arabic. The late successive BiTD group consisted of eight children, most of whom (seven out of eight) were recent arrivals from Syrian refugee backgrounds and started acquiring German upon school entry between ages 5;0 and 7;0. At the time of testing, all late successive BiTD children were enrolled in German mainstream schools for at least 2 years ($M = 2;2$, $SD = 4.76$ months). None of the participants in the late successive bilingual group had had formal schooling prior to arrival in Germany, and only two of them were enrolled in Sunday schools and had literacy skills in Arabic. Hence, the IL2 BiTD were comparable to their heritage peers in terms of Arabic schooling and literacy skills. As to the heritage BiSLI group, seven out of nine children were born in Germany (six simultaneous and one early successive). The remaining two children arrived with their families as refugees and at ages 4;2, and 4;8. Only one child in the BiSLI group received formal instruction in

¹⁵⁴ Two of these children migrated with their families as refugees.

Arabic and had rudimentary literacy skills in Arabic. The MoTD were, on average, younger and had a narrower age range than the bilingual groups ($M = 6;4$, age range: 5;1–8;7). The monolingual sample included both preschool children (14 out of 24) and primary school children. All of the monolingual school children ($n = 10$) had formal instruction in MSA and had Arabic literacy skills.

Table 5: Participants: an overview of background variables (Mean, SD, and range)

	MoTD (<i>n</i> = 24)	BiTD (<i>n</i> = 26)	BiSLI (<i>n</i> = 9)	
		Heritage BiTD	IL2 BiTD (<i>n</i> = 8)	
Age at testing (in months)	75.9 (11.8)	85.1 (12.5) 70-108	96.1 (6.81) 88-108	83.3 (13.7) 68-109
Age of Onset to L2 (AoO_L2 in months)	N/A	34.2 (10.2) 10-50	71.0 (4.84) 64-76	20.22 (23.3)
Length of Exposure to L2 (LoE_L2 in months)	N/A	50.9 (15.6) 20-74	25.9 (4.76) 19-33	63.1 (27.5) 19-99
%Early HL exposure (/100)	N/A	76.1 (20.1) 50-100	100 (0)	71.7 (20.1) 43-100
%Early L2 exposure (/100)	N/A	20.8 (19.6) 0-50	0 (0)	65.6 (38.7) 0-100
Current HL use (/16)	N/A	9.94 (2.86) 1-14	11.9 (2.79) 8-15	8.77 (4.58) 2-15
Current L2 use (/16)	N/A	5.22 (1.8) 2-8	4.13 (2.90) 1-8	6.89 (4.59) 1-14
HL richness (/14)	N/A	4.72 (2.42) 0-11	6.38 (1.77) 4-9	3.88 (1.62) 2-7
L2 richness (/14)	N/A	9.44 (2.43) 4-14	8.87 (2.53) 4-12	10.0 (2.55) 5-13
HL Literacy (yes/total)	(10/24)	(2/18)	(2/8)	(1/9)
Language Dominance Index, LDI*	N/A	-7.19 (9.02) -25.5-+11.0	-21.0 (8.0) -31.0- -	1.78 (18.4) -28-+24
Socioeconomic status (SES) (years of mother’s education)	15.0 (1.57)	14.1 (4.41) 8-26	15.0 (5.78) 7-22	13.7 (3.71) 7-20
Parental input quality (/16)	N/A	14.8 (2.33) 9-16	16.0 (0)	14.1 (2.2) 10-16

Notes. N/A = not applicable, MoTD = monolingual typically developing children; BiTD = bilingual children with typical language development; BiSLI = bilingual children with SLI; AoO = age of onset; L2 = second language; IL2 = late successive; LoE = length of exposure; HL = heritage language. *LDI<-5 = L1 dominant; -5-+5 = balanced; > +5 L2 dominant.

Comparisons of background variables related to bilingualism between the heritage BiTD, IL2 BiTD, and BiSLI groups revealed the following significant overall effects of Group: AoO_L2 [$\chi^2(2, N = 35) = 19.9, p < .001, \eta^2 = .561$], LoE_L2 [$\chi^2(2, N = 35) = 12.44, p = .002, \eta^2 = .326$], early L1 exposure [$\chi^2(2, N = 35) = 11.367, p = .003, \eta^2 = .293$], early L2 exposure [$\chi^2(2, N = 35) = 11.114, p = .004, \eta^2 = .285$], L1 richness [$\chi^2(2, N = 35) = 6.22, p = .045, \eta^2 = .132$], and current L1 skills [$\chi^2(2, N = 35) = 15.366, p < .001, \eta^2 = .418$]. The three groups were comparable in terms of current L1 use [$\chi^2(2, N = 35) = 3.73, p = .185, \eta^2 = .043$], current L2 use [$\chi^2(2, N = 35) = 2.382, p = .304, \eta^2 = .012$] and L2 richness [$\chi^2(2, N = 35) = 1.44, p = .487, \eta^2 = .018$]. Concerning parental input quality, only a marginally significant effect emerged [$\chi^2(2, N = 35) = 5.698, p = .058, \eta^2 = .116$]. Subsequent Mann-Whitney U tests applying Holm-Bonferroni adjustment showed no significant differences between heritage BiTDs and BiSLIs for any of the exposure variables. Interestingly, despite having similar exposure patterns to L1 and L2, a significant difference with large effect sizes emerged between heritage BiTDs and BiSLIs for the children's current L1 skills (oral abilities) as estimated by their parents ($U = 19.0, p = .006, r = .618$). Compared to their IL2 BiTD peers, heritage BiTDs started their systematic exposure to German earlier in childhood (AoO_L2: $U = 0.0, p < .001, r = .788$), had a considerably longer length of exposure to the L2 (LoE_L2: $U = 9.00, p < .001, r = .687$), and differed from IL2 BiTD children in the quality and quantity of early exposure to the L1 and L2 with IL2 children having considerably more early L1 exposure and less L2 exposure than their heritage peers (early L1 exposure: $U = 24.0, p = .018, r = .578$; early L2 exposure: $U = 28.0, p = .026, r = .541$). These differences in timing, length, and patterns of exposure to the L1/L2 resulted in a significant difference in the degree of language dominance between the latter two groups (LDI: $U = 19.0, p = .012, r = .578$), where IL2 BiTDs tended to show a higher degree of dominance in the L1. Notably, no significant differences were observed between heritage and IL2 BiTDs for current L1/L2 use (current L1 use: $U = 42.5, p = .102$; current L2 use: $U = 52.0, p = .285$), L1/L2 linguistic richness (L1 richness: $U = 41.0, p = .90$; L2 richness: $U = 63.0, p = .644$), or parental input quality ($U = 48.0, p = .071$). Moreover, heritage BiTDs did not differ from their IL2 BiTDs regarding current L1 skills according to parental assessment ($U = 44.5, p = .129$). A similar picture emerged for bilingualism-related variables when IL2 BiTDs were compared to the BiSLI group. IL2 BiTDs differed significantly from BiSLIs in AoO_L2 ($U = 0.0, p < .001, r = .846$), LoE_L2 ($U = 12.0, p = .021, r = .562$), early L1 and L2

exposure (early L1 exposure: $U = 8.0, p = .008, r = .736$; early L2 exposure: $U = 8.0, p = .006, r = .736$), and were considerably more dominant in the L1 than the BiSLI children ($U = 10.5, p = .028, r = .595$). Concerning input quality IL2 BiTDs significantly differed from the BiSLIs on the variables L1 richness ($U = 11.0, p = .030, r = .590$) and parental input quality ($U = 16.0, p = .017, r = .580$). On the other hand, IL2 BiTDs did not differ from BiSLIs with respect to current L1 and L2 use (current L1 use: $U = 42.5, p = .096$; current L2 use: $U = 52.0, p = .257$) or L2 linguistic richness ($U = 63.0, p = .612$). Similar to their heritage peers, IL2 BiTDs showed superior L1 skills (as estimated by the parents) to the BiSLI children (current L1 skills: $U = 1.5, p < .001, r = .812$).

4.5.2 Materials

Standardized test battery for Arabic (ELO-L)

For assessing oral proficiency in Arabic, the ELO-L comprehensive test battery for Lebanese Arabic (see **3.2.1.1** for a detailed description of the test) or dialectal adaptations thereof was administered to all bilingual¹⁵⁵ participants and 20 out of the 24 children of the monolingual control group. It is the first norm-referenced language assessment battery in Lebanon and one of the very few available standardized assessment tools for Spoken Arabic considering all dialects; see Shaalan (2017) and Rakhlin et al. (2021) for example tests for Gulf Arabic. The test comprises five subtests assessing receptive vocabulary (picture selection task), expressive vocabulary (picture naming task), morphosyntax comprehension (picture selection task), morphosyntax production (picture with sentence completion task), and phonology (word repetition task), and takes 30–45 min to administer. The Lebanese version of the ELO-L was adapted to other varieties of Arabic, including Levantine ones (Palestinian/Jordanian, Syrian), by the test authors in collaboration with linguistically trained native speakers of the respective varieties. During the adaptation process, special attention was paid to vocabulary and the phonological characteristics of the different Levantine Arabic varieties. It should, however, be noted that bilingual age-referenced norms only exist for the Lebanese variety covering the age range 3;0–7;11. Although the ELO-L takes the bilingual situation in Lebanon into

¹⁵⁵ The ELO-L was administered as part of the testing session assessing L1 abilities along with the Arabic LITMUS-SRT.

account, this kind of bilingual situation differs from what bilingual children experience in heritage contexts, where Arabic represents a minority language that is acquired in a nondiglossic context. Hence, for classifying the bilingual children into typically developing versus SLI based on the ELO-L, the test norms were adjusted according to the degree of language dominance of the respective children as estimated by the PaBiQ; that is, a less-stringent cutoff was chosen depending on whether the child was HL dominant (-1.5 SD), balanced (-1.75 SD), or L2 dominant (-2.25 SD), see 3.2 for details.

The Arabic LITMUS-SRT

This study uses the short version¹⁵⁶ of the Lebanese Arabic LITMUS-SRT (Zebib et al., 2020a). Similar to the German LITMUS-SRT (outlined in 3.3.1), the Arabic LITMUS-SRT was constructed according to the LITMUS principles (Marinis & Armon-Lotem, 2015) and incorporates complex structures cross-linguistically difficult for children with SLI in addition to simpler language-specific structures. The Lebanese version was adapted (phonologically and lexically) by the author of the present study into Palestinian/Jordanian and Syrian Arabic in collaboration with the task authors and linguistically trained native speakers of the respective dialects. The current version consists of 36 sentences varying in their structural (computational) complexity. As can be seen in **Table 6**, the primary complexity aspects targeted by the task are clausal embedding (presence of a subordinate clause) and *wh*-movement, that is, displacement of a syntactic constituent. Note that another complexity factor is involved in the object which-noun phrase (NP) questions, namely, the presence of a lexically restricted *wh*-element resulting in intervention effects (Friedmann et al., 2009). To ensure that lexical complexity does not act as a confounding factor, early acquired and highly frequent lexical items (e.g., boy, girl, mother, teacher, father, window) were deployed across different sentence types (Chevrie Muller & Maillart, 2009), controlling for their frequency throughout the task.

The following structure types are included in the task: simple (1) imperfective and (2) perfective sentences, (3) object *wh*-questions with an intervening NP (which-NP) alternating between resumptive clitic and gap strategies (Aoun et al., 2010), (4) finite and non-finite complement clauses, as well as definite (5) subject and (6) object relatives. In

¹⁵⁶ The original long version contains 64 sentences and was shortened to 36 sentences following a number of pilot studies conducted within the Lebanese context, e.g., Henry et al. (2012) and Henry and Saad (2012).

terms of length, the stimulus sentences contain three to six words, ranging from seven to 12 syllables per sentence. Wherever possible, VS word order alternated with SV word order, and subject–verb agreement alternated between masculine and feminine. All of the sentences in the Arabic LITMUS-SRT use singular determiner phrases that only require singular verbal agreement. Examples of the structures mentioned above are given in Table 7. The stimuli of the simple imperfective condition are of the type present progressive. Thus, they require using the proclitic marking the progressive aspect “ʕam” (see example 4.1 below) in addition to affixation. Therefore, they are structurally more complex than the past sentences included in the simple perfective condition.

Table 6: Arabic LITMUS-SRT: an overview of test conditions (adapted from Saad & Henry, 2020)

complexity factor		sentence type		number of sent.
Sentences without embedding (monoclausals)	-MOV	Simple perfective	Past	6
		Simple imperfective	Present	6
	+MOV to left periphery	Object- <i>wh</i> -question (which NP)	<i>wh</i> -question (object)	6
Sentences with embedding	-MOV	Finite complement clause	COMPL	3
		Non-finite complement clause		3
	+MOV to left periphery	Subject relative clause	SR	6
		Object relative clause	OR	6

increasing complexity



(4.1) Example simple imperfective sentence

عم ياكل الارنب جزرة

ʕam je:kol l'- arnab ʒazra
 PROG eat (3P sg masc impf) the-rabbit carrot (fem)
 ‘The rabbit is eating a carrot’

Concerning relative clauses, it should be noted that Arabic distinguishes between definite relatives¹⁵⁷ and indefinite ones, where the presence of a relativizer is conditioned

¹⁵⁷ According to Aoun (2000), movement is available in definite relatives but not in indefinite ones (p.14).

by the definiteness of the relativized noun (Aoun, 2000; Aoun & Choueiri, 1996; Galal, 2004). The relativizer *illi*¹⁵⁸ and its different cross-dialectal phonological manifestations (e.g., *jalli*, *halli*, *lli*, *balli*) “relativize only a definite head noun”. In contrast, no relativizer is required if the relativized noun is indefinite (Galal 2004, p. 46). Note that in MSA and most Arabic varieties, object relative clauses need an obligatory resumptive clitic (in the form of a verbal clitic in case of direct or indirect object relative clauses)¹⁵⁹. The resumptive clitic has to agree in person, gender, and number with its antecedent, as demonstrated in examples (4.2) and (4.3).

(4.2) Definite object relative:

شفت الولد يلي الرجال صورو

ʃefet elwalad jalle er-reʒʒe:l sawwar-o

see (1P sg perf) **the boy** (masc) **whom** the man (masc) photograph (3P sg perf) **-him**
(resumptive clitic)

‘I saw the boy whom the man photographed’

(4.3) Indefinite object relative:

شفت ولد الرجال صورو

ʃefet walad er-reʒʒe:l sawwar-o

see (1P sg perf) boy (masc) the man (masc) photograph (3P sg perf) **-him** (resumptive clitic)

‘I saw a boy whom the man photographed’

¹⁵⁸ Several studies on different Arabic dialects have treated *illi* as a (definite) complementizer in relative clauses, e.g., Jordanian Arabic (Al Momani, 2015), Syrian Arabic (Brustad, 2000), Moroccan Arabic (Benmamoun, 2000), Egyptian Arabic (Soltan, 2010), and Lebanese Arabic (Aoun, 2000). The latter studies view *illi* as a phonologically reduced equivalent of MSA’s relative complementizer *ʔallaḍi*, see Jarrah (2017) for a discussion. In addition, Shlonsky (1992) argues that in Palestinian Arabic *illi* only heads CPs functioning as predicates and “never heads subordinate clauses” (p.451). Based on Rizzi’s (1990) feature system, the C⁰ *illi* can be classified a [+predicational].

¹⁵⁹ Concerning relative clause formation in Levantine Arabic, Shlonsky (1992) argues that only subject relatives are generated via A’-movement, while object relatives featuring a resumptive clitic are formed through binding, where resumptive elements are employed as a “last resort strategy”. However, Aoun and Choueiri (1996) and Aoun et al. (2001) challenged Shlonsky’s account. They argued that, whenever possible, resumptive elements must be derived by movement (denoted as “apparent resumption”) and that binding is only employed when movement is blocked (see Botwinik et al., 2015 for a discussion).

Table 7: Structures included in the Arabic LITMUS-SRT with examples and glosses

Core structure	Substruct. (word order)	Example
(1) simple imperfective	SV	المعلمة عم تقرأ قصة l-emʕalme ʕam te'ra 'ossa the-teacher (fem) PROG read (3P sg fem impf) story (fem) 'The teacher is reading a story'
	VS	عم يعمل الصبي غاتو ʕam jaʕmol es-sabe gato PROG make (3P sg masc impf) the-boy (masc) cake (masc) 'The boy is making cake'
(2) simple perfective	SV	تيتا كسرت صحن te:ta: kasaret sahen grandma (fem) break (3P sg fem perf) plate (masc) 'Grandma broke a plate'
	VS	جابت المعلمة ورقة ʒe:bit l-emʕalme war'a bring (3P sg fem perf) the teacher (fem) paper (fem) 'The teacher brought a paper'
(3) object wh-questions	resumptive pronoun (SV)	أي كلب الحمار وقّعوا؟ ajja kaleb l-ehma:r wa'aʕ-o? which dog (masc) the donkey (masc) knock over (3P sg masc perf)-him 'Which dog did the donkey knock over?'
	gap (SV)	أي ولد الرجل صوّر؟ ajja walad er-reʒe:l sawwar? which boy (masc) the man (masc) photograph (3P sg masc perf) 'Which boy did the man photograph?'
(4) complement clauses	finite (VS)	فكر جدو انو الأرنب أكل موزة fakkar ʒeddo 'enno l-'arnab 'akal mawze (fem) think (3P sg masc perf) grandpa (masc) that the rabbit (masc) eat (3P sg masc perf) banana (fem) 'Grandpa thought that the rabbit ate a banana'
	non-finite (SV)	جدو يحب يسوق طائرة ʒeddo biheb jsu:' tijja:ra grandpa (masc) like (3P sg masc impf) drive (3P sg masc impf) plane (fem) 'Grandpa likes driving a plane'
(5) subject relatives (SR)	reversible	شفت البنت يلي باست المعلمة ʒefet il-benet jalle be:s-it le-mʕalme see (1P sg perf) the girl (fem) who kiss (3P sg perf) the teacher (fem) 'I saw the girl who kissed the teacher'
	irreversible	شفت البنت يلي جابت اللّعة ʒefet il-benet jalle ʒe:bit elleʕbe see (1P sg perf) the girl (fem) who bring (3P sg perf) the doll (fem) 'I saw the girl who brought the doll'
(6) object relatives (OR)	reversible (VS)	شفت البنت يلي جابتا ماما ʒefet el-benet jalle ʒe:bet-a ma:ma: see (1P sg perf) the girl (fem) whom bring (3P sg perf fem) her mom (fem) 'I saw the girl whom mom brought/gave birth to'
	irreversible (VS)	شفت الشّبّاك يلي فتحو الصبي ʒefet eʃ-febe:k jalle fatah-o is-sabe see (1P sg perf) the window (masc) which open (3P sg perf masc)-him the boy (masc) 'I saw the window which the boy opened'

The task takes 5–10 min to administer. The test sentences are prerecorded and presented in a pseudorandomized order via noise-canceling headphones. In this study, the task was scored using the measures “identical repetition” and “target structure.” Although “identical repetition” is the fastest and easiest scoring method, it might disadvantage bilingual children since any error might render the response incorrect. Scoring by target structure, on the other hand, measures the preservation of a given construction regardless of lexical and morphological errors not affecting the realization of the targeted construction. The requirements for the target structure to be scored as “correct” depend on the elicited structure. For instance, errors concerning definiteness (addition or omission of the definite article) are generally not penalized, as in (4.4). However, for relative clause structures to be scored as ‘correct’, the relativized noun must be marked for definiteness. Thus, if the definite article of the relativized noun were left out, the sentence would be scored as incorrect for both scoring measures, see (4.5).

(4.4) Target structure: Finite complement clause

فكر جدو انو الأرنب أكل موزة

fakkar ʒeddo 'enno l-'arnab 'akal mawze
 think (3P sg masc perf) grandpa (masc) that the rabbit (masc) eat (3P sg masc perf) banana (fem)
 ‘Grandpa thought that the rabbit ate a banana’

Child repetition:

فكر جدو انو الأرنب أكل الموزة

fakkar ʒeddo 'enno l-'arnab 'akal el-mawze
 think (3P sg masc perf) grandpa (masc) that the rabbit (masc) eat (3P sg masc perf) the banana (fem)
 ‘Grandpa thought that the rabbit ate the banana’

Morphological error: addition of the definite article to the object noun.

Score for identical repetition: 0, score for correct target structure: 1

(4.5) Target structure: Subject relative

شفت البنت يلي جابت اللعبة

ʃefet el-benet jalle ʒe:bit elleʃbe
 see (1P sg perf) the girl (fem) who bring (3P sg perf) the doll (fem)
 'I saw the girl who brought the doll'

Child repetition:

شفت بنت يلي جابت اللعبة

*ʃefet benet jalle ʒe:bit elleʃbe
 see (1P sg perf) **girl** (fem) who bring (3P sg perf) the doll (fem)
 'I saw a girl who brought the doll'

Morphological error: omission of the definite article on the relativized noun.

Score for identical repetition: 0, score for correct target structure: 0.

Standardized tests for assessment of nonverbal IQ and working memory

The German version of the Raven's Coloured Progressive Matrices test (Bulheller & Häcker, 2002) was used to obtain a rough estimate of the children's nonverbal cognition. The test was conducted in the child's dominant language, either in the L1 or the L2 testing session. Children's working memory capacity was assessed using the Forward Digit Span (FDS) and Backward Digit Span (BDS) subtests from the German version of the Wechsler Intelligence Scale for Children–Fourth Edition test battery (Petermann & Petermann, 2011). Forward digit recall is associated with verbal short-term memory, while backward digit recall is associated with verbal working memory. The latter tests were only administered to the bilingual participants during a separate session assessing L2 abilities; see 3.2.2.2 for details.

The parental questionnaire PaBiQ

The LITMUS-PaBiQ (Tuller, 2015, see 3.2.3.3 for details) was used to gather background information about the children and their families. In addition to general demographic variables, the PaBiQ incorporates questions about the quantity and quality of past and current language input/output, that is, exposure, in each of the child's languages, both within and outside the nuclear family. The latter information is synthesized into multiple indices (e.g., early HL/L2 exposure, current HL/L2 use, current HL/L2 richness) that allow the calculation of an experiential language dominance index. This index serves as

a baseline for adjusting the pathology cutoff criterion of monolingual norm-referenced tests according to the dominance status of the language being assessed. The PaBiQ also delivers information on the parents' educational levels as a proxy for the family's SES. In light of previous research, the following PaBiQ variables/indices were probed as potential factors explaining variance in performance on the ELO-L and the Arabic LITMUS-SRT:

- Chronological age
- SES (operationalized by years of maternal education)
- AoO of exposure to the L2 (AoO_L2)
- LoE to the L2 (LoE_L2)
- Early exposure ratio to the HL and L2 before age 4 years: the percentage of contexts in which the child is exposed to either language relative to the total number of possible contexts of exposure (exchanges with the mother, father, grandparents, nanny, siblings, other adults, nursery, and kindergarten)
- Current HL and L2 use at home: frequency of exchanges (ranging between never and always) between the child and immediate family members (mother, father, siblings, and other caregivers)
- Current HL/L2 richness: calculated based on how often the HL or L2 is currently used in extracurricular weekly activities and interactions outside the home environment, that is, print exposure and reading, audiovisual media, exchanges with playmates and family friends. This index also counts as an estimate of current language richness and represents an input quality variable.

In addition to the latter variables/indices, an estimate of parental HL input quality (out of 16 points) was computed for the present study given that parental characteristics may influence the quality of language input they provide, especially in the HL (see Chondrogianni & Daskalaki, 2023). This composite score comprises two types of information. The first includes the parents' self-rated proficiency in spoken and written Arabic on a scale from 0 to 4, where a maximum of 4 points can be allocated to each parent. The second parental input quality value is based on the generational status of the parents on a 4-point scale: 1 = non-Arab, 2 = second-generation migrant (the parent was born in an Arab family in Germany), 3 = first-generation child migrant (the parent was born and partially raised in an Arabic-speaking country but migrated to Germany as a child before the age of 12 years), and 4 = first-generation migrant (the parent was born

and raised in an Arabic-speaking country and migrated to Germany as an adolescent or adult, i.e., after the age of 12 years). A maximum of 4 points can be allocated to each parent, yielding 8 points in total.

4.5.3 Data analysis

All standardized tests and the Arabic-LITMUS-SRT and the PaBiQ were administered per test instructions outlined in Chapter 3, Section 3.2, and Section 4.3.2 of the current chapter. Responses were recorded with special dictaphones and were evaluated offline by two independent, linguistically trained raters. For the Arabic LITMUS-SRT, analyses were carried out on the percentage of correct responses for correct “identical repetition” and target structure. As to the ELO-L, data analyses were performed using both raw scores and standardized z-scores.

Statistical analyses were conducted using IBM SPSS 27 (IBM Corp., released 2020). Nonparametric tests were used for group comparisons due to unequal and small sample sizes and violation of normality assumptions, checked by the Shapiro–Wilk test. Hierarchical regression analyses were used to examine factors explaining variance in performance on the ELO-L subtests and the Arabic LITMUS-SRT. To avoid multicollinearity, the variance inflation factor (VIF) was checked to ensure that it fell within tolerable limits of a VIF value of 10. Only variables that yielded significant medium to strong correlations with the dependent measures were factored into the regression models as independent variables (Hair et al., 1995). The order of entry of the independent variables into the regression blocks was determined by the strength of the correlation between the dependent ELO-L or SRT performance measure and the independent variable. For the ELO-L, raw scores (the percentage of correct responses on each subtest) were considered for hierarchical regression analyses since they show more variability than standardized scores and since age was controlled for in the aforementioned analyses. Given the small sample size of the BiSLI group, regression analyses were limited to BiTD collapsing heritage and late successive BiTD together. A maximum of three blocks was entered into the hierarchical regression models with one independent variable per block due to the relatively small sample size ($n = 26$). This constellation is considered appropriate in light of recent statistical research (see Austin & Steyerberg, 2015; Field, 2018).

4.6 Results

4.6.1 Standardized test for Arabic (ELO-L)

For all subtests, analyses were performed on both raw scores (calculated as the total raw score divided by the maximum attainable score for each subtest multiplied by 100) and standardized z-scores. The latter scores control for age effects since they are calculated using the age means and standard deviations provided by the ELO-L test manual. For bilingual children older than the norming sample, z-scores were calculated relative to the norms of the oldest age range with available norms, that is, age range: 7;0–7;11 (cf. Hamann et al., 2020). **Tables 8** and **9** present descriptive statistics for the groups' overall performance on the different subtasks of the standardized Arabic test battery ELO-L. In **Table 8**, results for each subtest are reported as mean percentages of correct responses. In **Table 9**, results are given in standardized scores (z-scores), calculated relative to the corresponding age group per the test manual. A Kruskal-Wallis test revealed significant overall effects of Group on performance in all ELO-L subtests for both raw and z-scores (see **Tables 8** and **9**). Results for each subtest are presented in the subsequent sections, including descriptive statistics and the outcomes of the pairwise Mann–Whitney U comparisons with Holm–Bonferroni–adjusted *p* values.

Table 8: Group means (SD) on ELO-L subtests (raw scores, percentage correct)

ELO-L subtest	MoTD	IL2 BiTD	Heritage BiTD	BiSLI	Kruskal-Wallis test
Receptive vocabulary (%correct out of 34)	83.4% (10.7)	84.2% (11.3)	61.43% (15.2)	43.5% (13.3)	$\chi^2(3, N=52) = 31.2$ $p < .001$ $\eta^2 = .588$
Expressive vocabulary (% correct out of 70)	71.4% (16.7)	86.1% (5.1)	53.8% (17.6)	17.8% (25.3)	$\chi^2(3, N=52) = 28.9$ $p < .001$ $\eta^2 = .540$
Morphosyntax comp. (%correct out of 27)	70.6% (12.8)	72.2% (15.7)	64.2% (16.3)	35.8% (19.4)	$\chi^2(3, N=52) = 15.1$ $p = .002$ $\eta^2 = .252$
Morphosyntax production (%correct out of 37)	72.6% (14.9)	74.1% (8.4)	56.9% (16.6)	18.5% (19.8)	$\chi^2(3, N=52) = 25.7$ $p < .001$ $\eta^2 = .473$
Phonology (%correct out of 28)	88.6% (10.3)	94.6% (6.0)	85.3% (10.5)	37.8% (21.1)	$\chi^2(3, N=52) = 19.5$ $p < .001$ $\eta^2 = .344$

Table 9: Group means (SD) on ELO-L subtests (standardized z-scores)

ELO-L subtest	MoTD	IL2 BiTD	Heritage BiTD	BiSLI	Kruskal-Wallis test
Receptive vocabulary	1.32 (.601)	.511 (1.01)	-1.07 (1.35)	-2.7 (1.2)	$\chi^2(3, N=52) = 35.2$ $p < .001$ $\eta^2 = .671$
Expressive vocabulary	7.26 (1.10)	.723 (.532)	-1.93 (2.01)	-5.51 (2.6)	$\chi^2(3, N=52) = 33.2$ $p < .001$ $\eta^2 = .629$
Morphosyntax comprehension	3.10 (2.62)	.738 (1.51)	.521 (1.34)	-1.94 (1.6)	$\chi^2(3, N=52) = 24.4$ $p < .001$ $\eta^2 = .446$
Morphosyntax production	1.43 (1.63)	-.246 (.842)	-.742 (1.44)	-3.6 (1.82)	$\chi^2(3, N=52) = 28.8$ $p < .001$ $\eta^2 = .538$
Phonology	-.370 (1.45)	-.536 (1.36)	-1.36 (1.56)	-8.41 (3.8)	$\chi^2(3, N=52) = 20.6$ $p < .001$ $\eta^2 = .367$

4.6.1.1 Lexical abilities

As can be seen in **Table 10**, the heritage BiTD group scored significantly lower on expressive and receptive vocabulary than the MoTD and IL2 BiTD groups, but higher than the BiSLI group. Although the heritage BiTD group showed superior lexical abilities to their counterparts with SLI, 12/18 heritage BiTDs would be classified as BiSLI when compared to the norms provided by the test manual, despite adjusting the cutoff points in accordance with their language dominance. In contrast, no significant differences were found between the MoTD and IL2 BiTD groups for either lexical measure, and except for one late successive child, all MoTD and IL2 BiTD children would be classified as typically developing based on the test. **Figures 6** and **7** depict the groups' overall performance on the ELO-L subtests assessing receptive and expressive vocabulary.

Table 10: Pairwise Mann-Whitney U comparisons: ELO-L (percentage correct and z-scores), subtests Receptive (LexR) and Expressive vocabulary (LexP)

	LexR %correct	LexR z-score	LexP %correct	LexP z-score
MoTD vs. BiSLI	$U = 0.00$ $p < .001$ $r = .798$	$U = 0.00$ $p < .001$ $r = .796$	$U = 9.00$ $p < .001$ $r = .709$	$U = 2.00$ $p < .001$ $r = .770$
Heritage BiTD vs. BiSLI	$U = 33.0$ $p = .039$ $r = .477$	$U = 31.5$ $p = .044$ $r = .490$	$U = 20.0$ $p = .012$ $r = .605$	$U = 22.5$ $p = .015$ $r = .579$
IL2 BiTD vs. BiSLI	$U = 2.00$ $p < .001$ $r = .794$	$U = 3.00$ $p = .006$ $r = .771$	$U = 0.00$ $p < .001$ $r = .842$	$U = 0.00$ $p < .001$ $r = .841$
MoTD vs. Heritage BiTD	$U = 42.0$ $p < .001$ $r = .648$	$U = 18.0$ $p < .001$ $r = .762$	$U = 85.0$ $p = .030$ $r = .457$	$U = 40.0$ $p < .001$ $r = .673$
MoTD vs. IL2 BiTD	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Heritage BiTD vs. IL2 BiTD	$U = 14.5$ $p = .006$ $r = .620$	$U = 23.5$ $p = .021$ $r = .519$	$U = 9.50$ $p < .001$ $r = .670$	$U = 13.0$ $p < .001$ $r = .631$

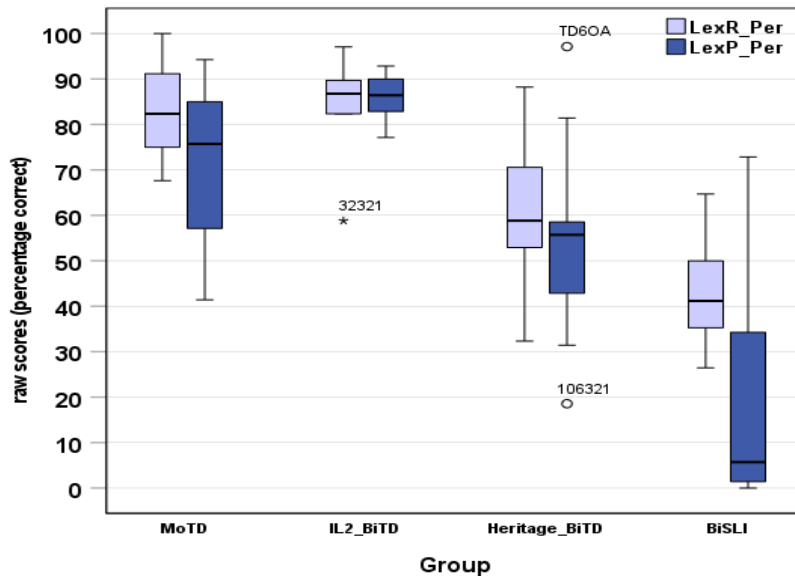


Figure 6: ELO-L (raw scores, %correct): subtests Receptive (LexR) and Expressive Vocabulary (LexP)

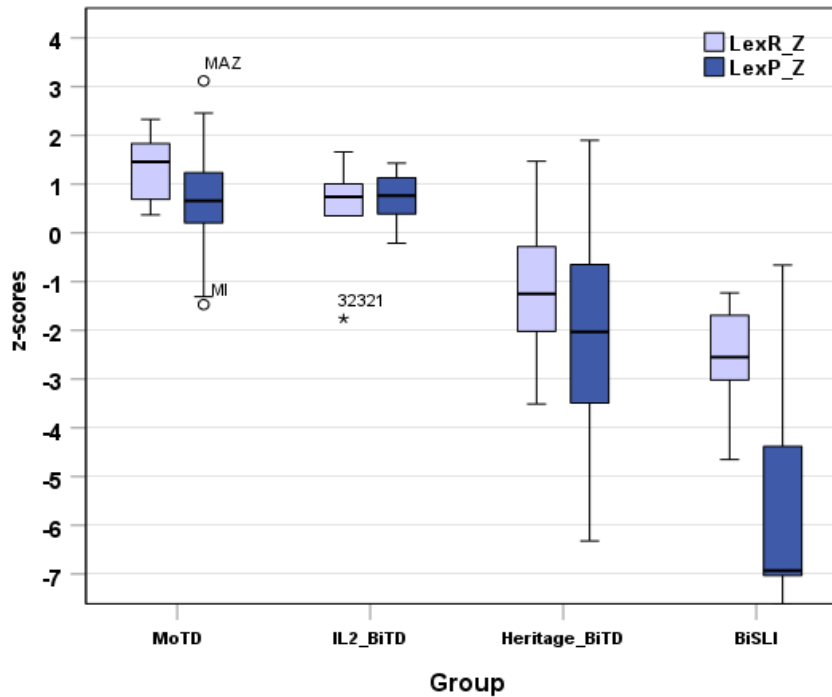


Figure 7: ELO-L (z-scores): subtests Receptive (LexR) and Expressive Vocabulary (LexP)

4.6.1.2 Morphosyntactic abilities

Regarding HL morphosyntax production (see **Figures 8 and 9**), the performance of the BiSLI group was significantly worse than the three TD groups, regardless of whether raw scores or z-scores were considered. Furthermore, the heritage BiTDs were outperformed by the monolingual and late successive groups when raw scores were considered. However, when z-scores were considered, a significant difference emerged between MoTDs and heritage BiTDs, as well as late successive BiTDs, but not between heritage BiTDs and late successive BiTDs. Based on the norm references, 9/18 heritage BiTDs would count as SLI, while all late successive BiTDs would be classified as typically developing. As to morphosyntax comprehension, all three typically developing groups showed superior performance to heritage BiSLIs for both raw and z-scores. When raw scores were considered, heritage BiTDs performed on par with MoTDs and late successive BiTDs. As to z-scores, significant differences emerged between the MoTD and both BiTD groups, while no significant differences emerged between heritage BiTDs and late successive BiTDs. Interestingly, only one heritage BiTD and one IL2 BiTD would be considered BiSLI based on morphosyntax comprehension. For a summary of pairwise comparisons, see **Table 11**.

Table 11: Pairwise Mann-Whitney U comparisons: ELO-L (percentage correct and z-scores), subtests Morphosyntax Comprehension (MorphR) and Morphosyntax Production (MorphP)

	MorphR %correct	MorphR z-score	MorphP %correct	MorphP z-score
MoTD vs. BiDLD	$U = 17.0$ $p < .001$ $r = .641$	$U = 9.00$ $p < .001$ $r = .712$	$U = 3.50$ $p < .001$ $r = .738$	$U = 0.00$ $p < .001$ $r = .769$
Heritage BiTD vs. BiDLD	$U = 18.0$ $p = .006$ $r = .625$	$U = 22.5$ $p = .012$ $r = .579$	$U = 11.5$ $p = .006$ $r = .660$	$U = 17.5$ $p = .010$ $r = .594$
IL2 BiTD vs. BiDLD	$U = 7.00$ $p = .030$ $r = .679$	$U = 13.0$ $p = .027$ $r = .537$	$U = 0.00$ $p = .006$ $r = .841$	$U = 1.00$ $p = .006$ $r = .814$
MoTD vs. Heritage BiTD	$n.s.$	$U = 70.0$ $p = .006$ $r = .505$	$U = 80.0$ $p = .012$ $r = .475$	$U = 57.5$ $p < .001$ $r = .581$
MoTD vs. IL2 BiTD	$n.s.$	$U = 18.0$ $p = .012$ $r = .594$	$n.s.$	$U = 27.5$ $p = .040$ $r = .505$
Heritage BiTD vs. IL2 BiTD	$n.s.$	$n.s.$	$U = 24.5$ $p = .024$ $r = .519$	$n.s.$

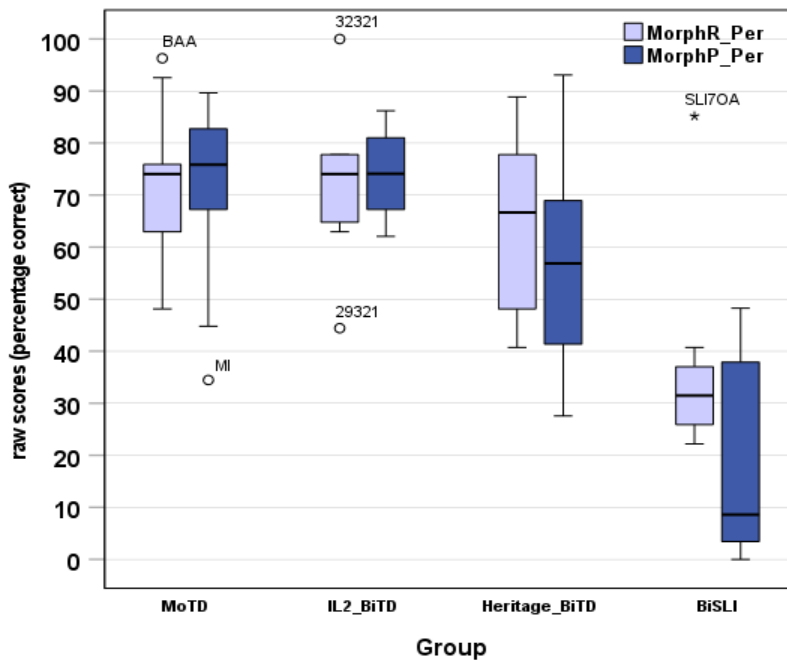


Figure 8: ELO-L (raw scores, %correct): subtests Morphosyntax Comprehension (MorphR) and Production (MorphP)

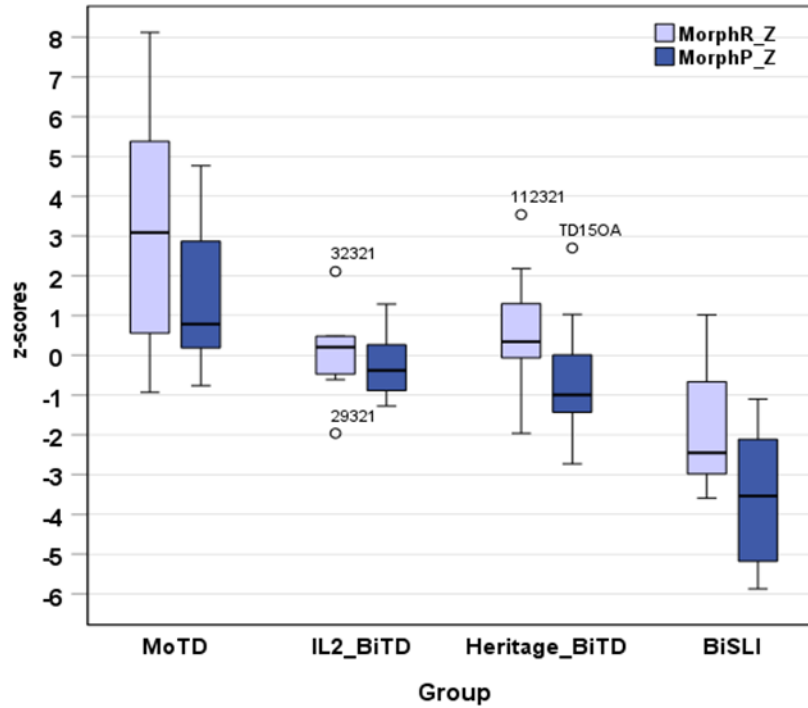


Figure 9: ELO-L (z-scores): subtests morphosyntax comprehension (MorphR) and production (MorphP)

4.6.1.3 Phonological abilities

Concerning phonology (see **Figures 10 and 11**), all typically developing groups performed significantly better than the BiSLI group with large effect sizes: MoTD vs. BiSLI: (raw scores: $U = 2.50$, $p < .001$, $r = .757$; z-scores: $U = 1.00$, $p < .001$, $r = .776$), heritage BiTD vs. BiSLI: (raw scores: $U = 5.00$, $p < .001$, $r = .704$; z-scores: $U = 2.00$, $p < .001$, $r = .738$), and IL2 BiTD vs. BiSLI: (raw scores: $U = 0.00$, $p < .001$, $r = .843$; z-scores: $U = 0.00$, $p < .001$, $r = .830$). No significant differences emerged between MoTD, heritage, and late successive BiTD. However, compared to the norms, 8/18 heritage BiTDs, 2/8 late successive BiTDs, and three preschool monolingual children (age < 6;0) would be classified as SLI based on phonology.

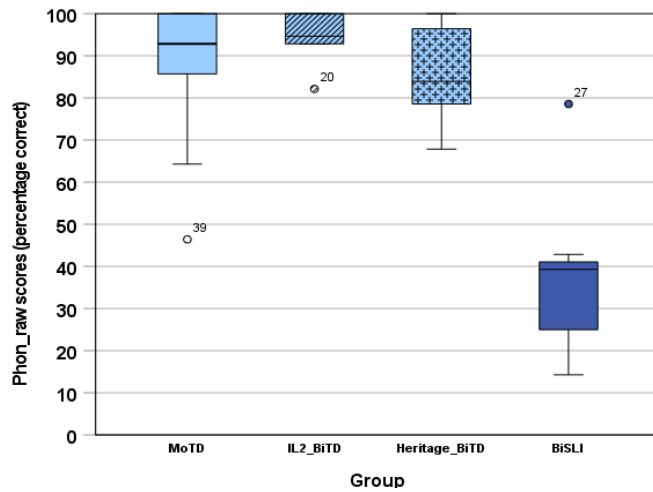


Figure 10: ELO-L: subtest Phonology (raw scores, %correct)

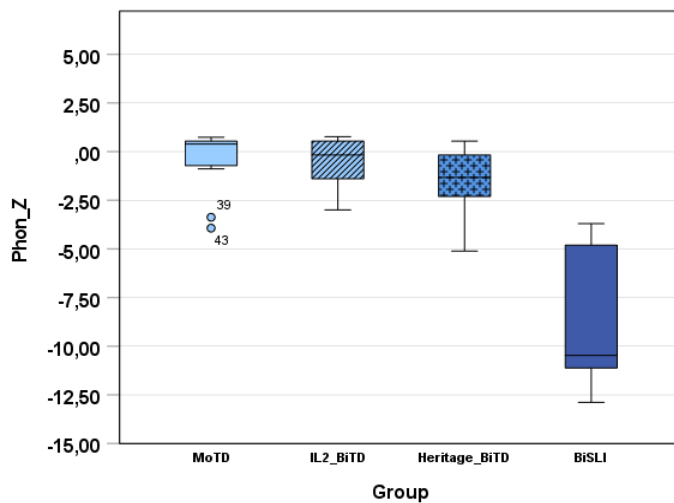


Figure 11: ELO-L: subtest Phonology (z-scores)

When the test battery is considered as a whole, 12/18 heritage BiTD children would fulfill the criteria for classification as SLI based on standardized assessment, that is, having at least two scores in the pathology range. In contrast, none of the monolingual or late successive BiTDs would be overdiagnosed with SLI, and all heritage BiSLI children had more than two scores in the pathology range with a tendency to floor performance. Taken together, while all BiSLI were identified as such by the test, a significant proportion of the heritage BiTD children were overidentified as having SLI, i.e., risk of overdiagnosis. Note that most of the heritage BiTDs were either balanced bilinguals or Arabic-dominant based on the PaBiQ estimates.

4.6.1.4 Predictors of Performance on the ELO-L Test Battery

To examine which age and input variables explained the variance in performance of the BiTD children on the ELO-L, bivariate Spearman correlations were first carried out within the age and input variables and performance scores on the different subtests of the ELO-L. Subsequently, hierarchical regression analyses were run using ELO-subtests as dependent variables and background variables, yielding moderate to strong correlations with the respective ELO-L subtests as independent variables (see **Table 12**). Since chronological age was part of the correlational matrix, modeling was done only with raw performance scores (converted to percentages) on the ELO-L. Although language dominance yielded significant correlations with most ELO-L subtests, it was not used as an independent variable in the hierarchical regression models since all of the variables related to bilingualism contribute to calculating the language dominance index. Considering sub-indices such as current language use, richness, or early exposure to the L1 and L2 allows for a more fine-grained analysis (see Tuller et al., 2018).

Table 12: BiTD children: Spearman correlations between ELO-L subtests and PaBiQ age and input variables

Age and input variables(PaBiQ)	ELO-L				
	LexR	LexP	MorphR	MorphP	Phon
Chronological age	.368	.316	.247	.147	.503**
Socioeconomic status (SES)	.196	-.011	.117	.155	.076
AoO_L2	.645**	.643**	.116	.504**	.303
LoE_L2	-.456*	-.573**	-.003	-.424*	-.102
Early HL exposure	.560**	.700**	.233	.500**	.318
Early L2 exposure	-.658**	-.702**	-.240	-.515**	-.411*
Current HL use	.352	.481**	.032	.400*	-.068
Current L2 use	-.457*	-.487*	-.138	-.445*	.213
HL richness	.402*	.462*	.221	.242	.008
L2 richness	-.125	-.227	-.182	-.147	.393
Parental HL input quality	.451*	.497*	-.179	.304	.052

Notes: Heritage and IL2 BiTD children were collapsed together for correlational analyses. LexR (receptive vocabulary), LexP (expressive vocabulary), MorphR (morphosyntax comprehension), MorphP (morphosyntax production), Phon (phonology). * $p < .05$, ** $p < .01$, *** $p < .001$. Significant correlations are given in bold.

Receptive and expressive vocabularies were significantly correlated with AoO and LoE to the L2, early HL and L2 exposure, current HL/L2 use, HL richness, and parental HL input quality (see **Table 12**). Due to collinearity issues between AoO_L2 and early L2 exposure, as well as between early HL exposure and early L2 exposure (as indicated by the VIF values), models treating both sets of variables as independent variables were disregarded. Concerning receptive vocabulary, the most parsimonious models (see **Table 13**) showed that AoO to the L2 and early exposure to the L2 explained 30.7% and 31.5% of the variance when entered into the first block of Models 1 and 2, respectively, with current L2 use explaining an additional 10.3% in Model 1 and 11.3% in Model 2. Parental input quality, on the other hand, did not have any significant contribution when entered in the last step of Model 1 and Model 2. Expressive vocabulary, on the other hand, was explained by a combination of early L2 exposure, explaining 43.5% of the variance in Step 1, and parental HL input quality, explaining an additional 14.9% of the variance in Step 2 of Model 1, and by AoO to the L2 as the sole significant predictor explaining 52.9% of the variance in Model 2 (see **Table 14**).

Table 13: Summary of hierarchical regression analysis for age and input variables predicting performance in the Receptive Vocabulary subtest of the ELO-L

		Model 1				Model 2		
		<i>b</i>	<i>SE b</i>	β		<i>b</i>	<i>SE b</i>	β
Step 1								
	Constant	45.7	7.55		Constant	76.65	3.813	
	AoO_L2	0.499	0.153	.554**	Early L2 exposure	-43.08	12.96	-.561**
Step 2								
	Constant	63.1	11.22		Constant	89.108	6.84	
	AoO_L2	0.410	0.151	.455*	Early L2 exposure	-36.16	12.54	-.471**
	Current L2 use	.311	.168	.314 <i>p</i> = .06	Current L2 use	-2.819	1.325	-.348*
Step 3								
	Constant	41.4	28.2		Constant	-56.89	26.58	
	AoO_L2	.398	0.153	.442*	Early L2 exposure	-36.91	12.39	-.481**
	Current L2 use	-2.243	1.484	-.277 <i>p</i> = .06	Current L2 use	-2.064	1.441	-.254
	Parental input quality	1.305	1.563	.149	Parental input quality	1.887	1.506	.215

Notes: Model 1: $R^2 = .307$ for Step 1: $\Delta R^2 = .103$ for Step 2 ($p = .057$): $\Delta R^2 = .018$ for Step 3 ($p = .413$).

Model 2: $R^2 = .315$ for Step 1: $\Delta R^2 = .113$ for Step 2 ($p < .05$): $\Delta R^2 = .012$ for Step 3 ($p = .492$).

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 14: Summary of hierarchical regression analysis for age and input variables predicting performance in the Expressive Vocabulary subtest of the ELO-L

		Model 1			Model 2			
		<i>b</i>	<i>SE b</i>	<i>β</i>		<i>b</i>	<i>SE b</i>	<i>β</i>
Step 1								
	Constant	74.377	4.040		Constant	27.683	7.533	
	Early L2 exposure	-73.782	17.181	-.659***	AoO_L2	.792	22.087	.727***
Step 2								
	Constant	12.388	21.897		Constant	-6.877	22.087	
	Early L2 exposure	-75.054	15.068	-.671***	AoO_L2	0.742	0.151	.681***
	Parental input quality	4.092	1.427	.386**	Parental input quality	2.426	1.464	.229
Step 3								
	Constant	29.681	28.049		Constant	17.046	28.234	
	Early L2 exposure	-70.324	15.819	-.628***	AoO_L2	0.691	0.152	.637***
	Parental input quality	3.394	1.593	.320*	Parental input quality	1.627	1.561	.154
	Current L2 use	-1.508	1.528	-.154	Current L2 use	-1.967	1.482	-.200

Notes: Model 1: $R^2 = .435$ for Step 1: $\Delta R^2 = .149$ for Step 2 ($p = .009$): $\Delta R^2 = .018$ for Step 3 ($p = .334$). Model 2: $R^2 = .529$ for Step 1: $\Delta R^2 = .050$ for Step 2 ($p = .111$): $\Delta R^2 = .031$ for Step 3 ($p = .198$). * $p < .05$, ** $p < .01$, *** $p < .001$

Regarding morphosyntax comprehension, no significant correlations were found between performance and any of the age and input variables; hence, no regression analyses were undertaken for this measure. On the other hand, morphosyntax production (see **Table 15**) was significantly correlated with AoO to the L2, LoE to the L2, early HL/L2 exposure, and current HL/L2 use. Early HL exposure accounted for 36.4% of the variance, with current L2 use explaining an additional 9.3% (Model 1). In Model 2, early L2 exposure explained 23.1% of the variance, with current L2 use adding 16.5%. In Model 3, AoO_L2 explained 22.5% of the variance, with current L2 use accounting for further 15.6% of the variance. Current HL use did not have a significant contribution to any of the variables. As to phonology (see **Table 16**), performance was only accounted for by chronological age, which explained 26.7% of the variance.

Table 15: Summary of hierarchical regression analysis for age and input variables predicting performance in the Morphosyntax Production subtest of the ELO-L

	Model 1				Model 2			
		<i>b</i>	<i>SE b</i>	β		<i>b</i>	<i>SE b</i>	β
Step 1								
	Constant	20.231	11.573		Constant	68.245	3.669	
	Early HL exposure	50.064	13.472	.604*	Early L2 exposure	-41.898	15.603	-.481*
Step 2								
	Constant	41.723	15.360		Constant	82.562	6.611	
	Early HL exposure	39.256	13.838	.474**	Early L2 exposure	-32.279	14.640	-.370*
	Current L2 use	-2.530	1.277	-.331 <i>p</i> = .060	Current L2 use	-3.215	1.284	-.421*
Step 3								
	Constant	52.657	26.787		Constant	92.572	23.166	
	Early HL exposure	38.867	14.089	.463*	Early L2 exposure	-31.750	14.946	-.364*
	Current L2 use	-3.231	1.905	-.423	Current L2 use	-3.880	1.969	-.506 <i>p</i> = .061
	Current HL use	-0.682	1.357	-.119	Current HL use	-0.649	1.437	-.133
	Model 3							
		<i>b</i>	<i>SE b</i>	β				
Step 1								
	Constant	43.901	7.525					
	AoO L2	0.402	0.153	.474*				
Step 2								
	Constant	64.075	10.827					
	AoO L2	0.298	0.146	.352*				
	Current L2 use	-3.163	1.312	-.414*				
Step 3								
	Constant	87.726	23.073					
	AoO L2	0.340	0.149	.401*				
	Current L2 use	-4.751	1.891	-.622*				
	Current HL use	-1.688	1.457	-.295				

Notes: Model 1: $R^2 = .364$ for Step 1: $\Delta R^2 = .093$ for Step 2 ($p = .060$): $\Delta R^2 = .006$ for Step 3 ($p = .620$).

Model 2: $R^2 = .231$ for Step 1: $\Delta R^2 = .165$ for Step 2 ($p = .020$): $\Delta R^2 = .006$ for Step 3 ($p = .656$).

Model 3: $R^2 = .225$ for Step 1: $\Delta R^2 = .156$ for Step 2 ($p = .024$): $\Delta R^2 = .036$ for Step 3 ($p = .259$).

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 16: Summary of hierarchical regression analysis for age and input variables predicting performance in the Phonology subtest of the ELO-L

Model 1				
		<i>B</i>	<i>SE b</i>	<i>β</i>
<i>Step 1</i>				
	Constant	49.44	13.216	
	Age	0.438	0.148	.517**
<i>Step 2</i>				
	Constant	63.094	14.629	
	Age	0.316	0.156	.455*
	Early L2 exposure	-15.181	8.255	-.338 <i>p</i> = .079

Notes: Model 1: $R^2 = .267$ for Step 1: $\Delta R^2 = .094$ for Step 2 ($p = .079$)

* $p < .05$, ** $p < .01$, *** $p < .001$.

4.6.2 Arabic LITMUS-SRT

As can be seen in **Figure 12**, unlike most of the ELO-L subtests, heritage bilingual children were not disadvantaged by the Arabic LITMUS-SRT, especially when the task was scored by target structure. Although heritage BiTD showed more variance than MoTD and late successive BiTD on “identical repetition” (heritage BiTD: $M = 66.5$, $SD = 20.7$; IL2 BiTD: $M = 80.6$, $SD = 16.9$; MoTD: $M = 81.0$, $SD = 13.3$) and target structure (heritage BiTD: $M = 76.3$, $SD = 17.8$; IL2 BiTD: $M = 89.9$, $SD = 11.5$; MoTD: $M = 88.5$, $SD = 10.3$), no significant differences emerged between heritage BiTD and either MoTD or late successive BiTD. Moreover, no significant differences were found between MoTD and IL2 BiTD, and all TD groups performed significantly better than the heritage BiSLI group with large effect sizes regardless of whether the task was scored by “identical repetition” or target structure: “identical repetition” (MoTD vs. BiSLI: $U = 1.5$, $p < .001$, $r = .619$; heritage BiTD vs. BiSLI: $U = 9.5$, $p < .05$, $r = .548$; IL2 BiTD vs. BiSLI: $U = 2.0$, $p < .05$, $r = .730$); target structure (MoTD vs. BiSLI: $U = 3.00$, $p < .001$, $r = .613$; heritage BiTD vs. BiSLI: $U = 10.5$, $p < .05$, $r = .528$; IL2 BiTD vs. BiSLI: $U = 1.00$, $p < .05$, $r = .776$).

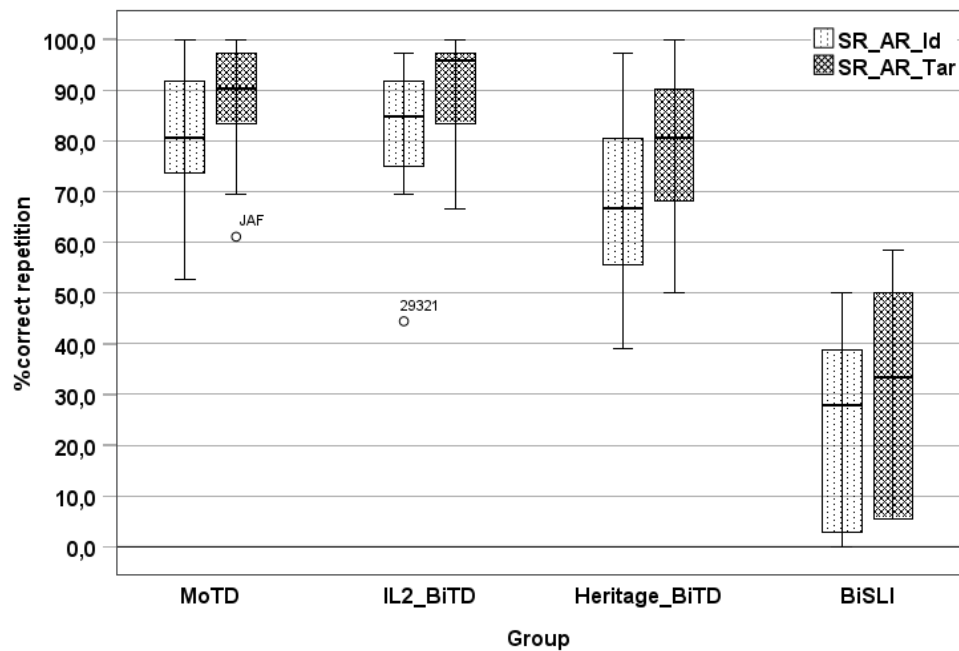


Figure 12: Arabic LITMUS-SRT: % correct identical repetition & target structure

4.6.2.1 Predictors of performance on Arabic LITMUS-SRT

No significant correlations emerged between performance on Arabic SRT and age and input variables (see **Table 17**). Hence, age and input variables were not examined in regression analyses. With respect to working memory and linguistic measures (subtests of the ELO-L), significant correlations emerged between “identical repetition” and HL morphosyntax comprehension, HL expressive vocabulary, HL phonology, and verbal short-term memory (indexed by FDS).

The regression analysis (see **Table 18**) showed that performance on “identical repetition” was mainly predicted by HL morphosyntax reception, accounting for 49.5% of the variance, and FDS explaining an additional 18.4%. Performance on target structure was correlated with FDS, HL morphosyntax comprehension, HL expressive vocabulary, and HL phonology in addition to HL morphosyntax production. Similar to “identical repetition,” performance on target structure was primarily predicted by morphosyntax reception, which explained 42.8% of the variance, with FDS explaining an additional 15.6%. For both measures, “identical repetition” and “target structure”, adding HL expressive vocabulary, HL phonology, or HL morphosyntax production in the third step to the models did not significantly contribute to explaining variance.

Table 17: BiTD children: Spearman correlations between Arabic LITMUS-SRT and PaBiQ age and input variables, as well as working memory and linguistic measures

	SRT_AR_Id	SRT_AR_Tar
Forward Digit Span (FDS, VSTM)	.537*	.465*
Backward Digit Span (BDS, VWM)	.134	.175
Language Variables (%correct, raw scores)		
LexR_ELO-L	.384	.400
LexP_ELO-L	.481*	.525*
MorphR_ELO-L	.581**	.629**
MorphP_ELO-L	.336	.440*
Phon_ELO-L	.535**	.456*
Age and input variables		
Chronological age	.157	.176
Socioeconomic status (SES)	.363	.326
AoO_L2	.166	.183
LoE_L2	-.237	-.276
Early HL exposure	.186	.120
Early L2 exposure	-.206	-.191
Current HL use	.150	.216
Current L2 use	.049	-.022
HL richness	.349	.386
L2 richness	.041	-.005
Quality of parental HL input	.002	.037

Notes: Heritage and IL2 participants were collapsed together for correlational analysis. SRT_AR_Id, Arabic LITMUS-SRT “identical repetition.” SRT_AR_Tar, Arabic LITMUS-SRT “target structure.”

* $p < .05$. ** $p < .01$. *** $p < .001$. Significant correlations are given in bold.

Table 18: BiTD children: Summary of hierarchical regression analysis for working memory and linguistic variables

	SRT_AR_Id				SRT_AR_Tar			
		<i>b</i>	<i>SE b</i>	β		<i>b</i>	<i>SE b</i>	β
Step 1								
	Constant	22.163	12.813		Constant	41.132	11.723	
	MorphR	.788	.187	.706**	MorphR	.627	.171	.654**
Step 2	Constant	-9.961	30.439		Constant	40.129	12.559	
	MorphR	.703	.199	.629**	MorphR	.598	.203	.624**
	Phonology	.417	.359	.207	LexP	.043	.154	.059
Step 3	Constant	-29.933	25.176		Constant	3.449	13.953	
	MorphR	.543	.167	.486**	MorphR	.329	.172	.343*
	Phonology	.243	.293	.120	LexP	.148	.121	.203
	FDS	9.103	2.817	.470**	FDS	9.332	2.577	.562**

Notes: SRT_AR_Id: $R^2 = .498$ for Step 1: $\Delta R^2 = .037$ for Step 2 ($p = .262$); $\Delta R^2 = .184$ for Step 3 ($p = .005$). SRT_AR_Tar: $R^2 = .428$ for Step 1: $\Delta R^2 = .003$ for Step 2 ($p = .783$); $\Delta R^2 = .256$ for Step 3 ($p = .002$). * $p < .05$, ** $p < .01$, *** $p < .001$.

4.6.2.2 Performance on the individual structures of the Arabic LITMUS-SRT

As to “identical repetition” (see **Figure 13** and **Table 19**), no significant differences were found between the typically developing monolingual and bilingual groups on any of the structures. Importantly, BiSLIs scored significantly worse than MoTDs and heritage BiTDs on all structures. Late successive BiTDs also significantly outperformed BiSLIs on all structures except for finite complement clauses. As to “target structure” (see **Figure 14**), heritage BiTDs performed on par with monolinguals and late successive BiTDs on all structures except for object relative clauses. At the same time, while heritage BiTDs significantly outperformed BiSLIs on most structure types, no significant differences emerged between heritage BiTDs and BiSLIs on object relatives (see **Table 20**).

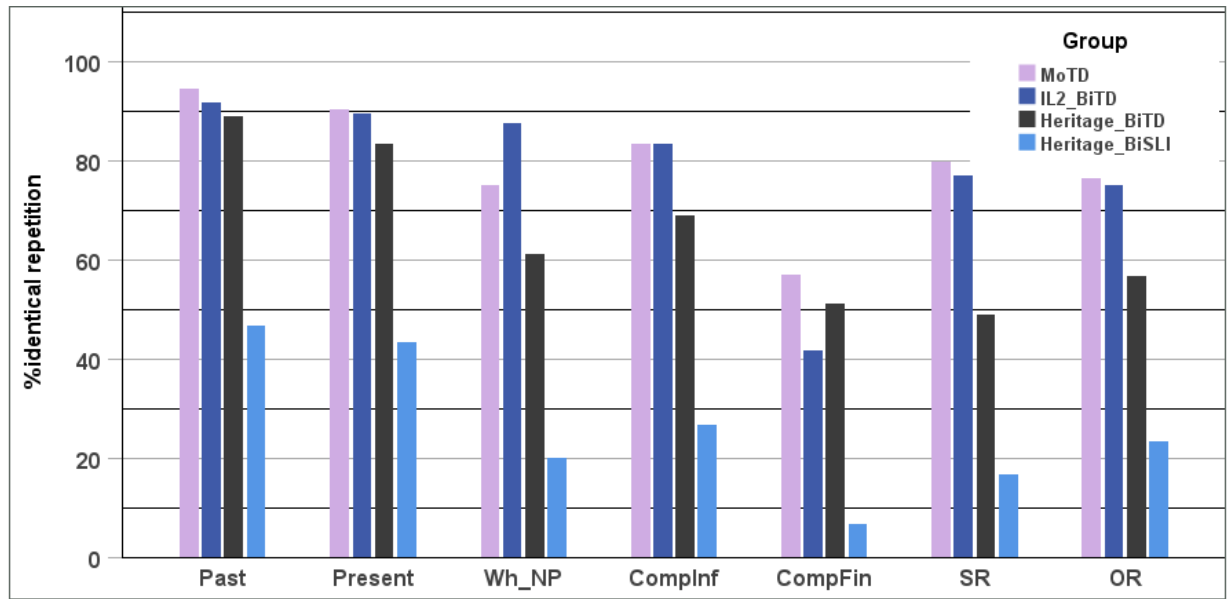


Figure 13: Arabic LITMUS-SRT: mean percentage of identical repetition on each structure subtype

Table 19: The effect of group membership on percentage identical repetition on each syntactic structure (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests)

[illegible]

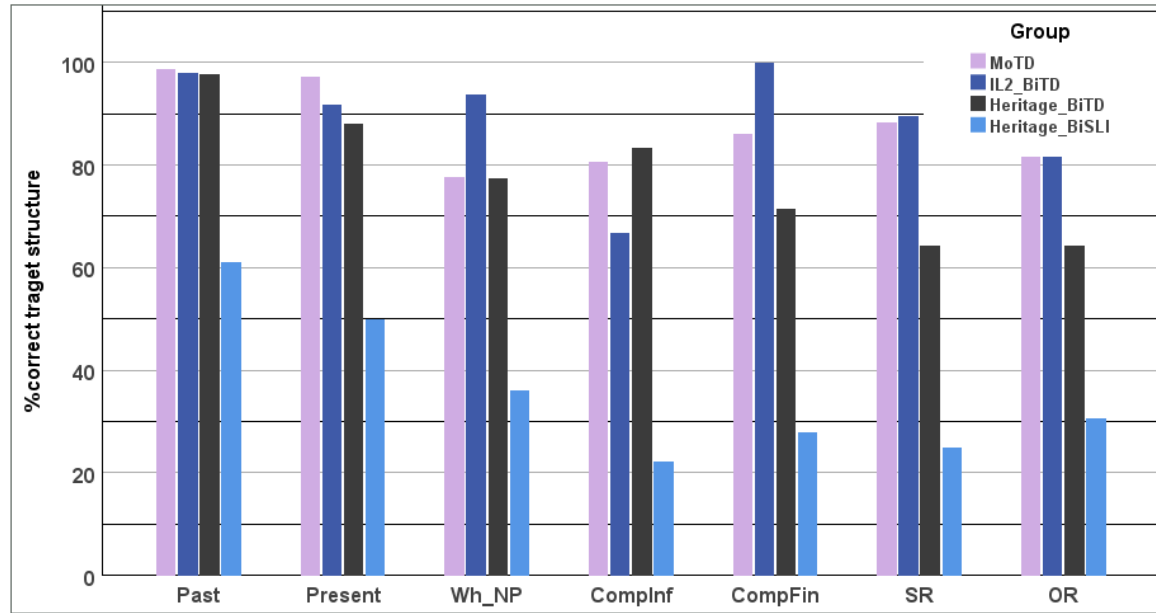


Figure 14: Arabic LITMUS-SRT: Mean percentage of correct target structure on each structure subtype

Table 20: The effect of group membership on percentage correct target structure on each syntactic structure (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests)

SRT_AR_Tar	Past	Present	Wh_NP	Complnf	CompFin	SR	OR
	$\chi^2(3, N=52) = 22.7$ $p < .001$	$\chi^2(3, N=52) = 20.4$ $p < .001$	$\chi^2(3, N=52) = 13.8$ $p = .003$	$\chi^2(3, N=52) = 9.38$ $p = .025$	$\chi^2(3, N=52) = 15.0$ $p = .002$	$\chi^2(3, N=52) = 17.8$ $p < .001$	$\chi^2(5, N=52) = 15.3$ $p = .002$
MoTD vs. BiSLI	$U = 14.0$ $p < .001$ $r = .741$	$U = 4.00$ $p < .001$ $r = .772$	$U = 20.5$ $p = .018$ $r = .501$	$U = 21.0$ $p = .012$ $r = .521$	$U = 22.0$ $p = .012$ $r = .550$	$U = 8.00$ $p < .001$ $r = .642$	$U = 19.5$ $p = .020$ $r = .520$
Heritage BiTD vs. BiSLI	$U = 9.5$ $p = .007$ $r = .741$	$U = 12.0$ $p = .040$ $r = .579$	$U = 11.0$ $p = .045$ $r = .584$	$U = 12.5$ $p = .042$ $r = .602$	$U = 18.5$ $p = .042$ $r = .454$	$U = 13.0$ $p = .045$ $r = .579$	<i>n.s.</i>
IL2 BiTD vs. BiSLI	$U = 5.00$ $p = .018$ $r = .728$	$U = 4.00$ $p = .024$ $r = .740$	$U = 1.50$ $p = .012$ $r = .811$	$U = 6.50$ $p = .045$ $r = .596$	$U = 4.00$ $p = .012$ $r = .810$	$U = 2.50$ $p = .020$ $r = .762$	$U = 7.00$ $p = .025$ $r = .598$
MoTD vs. Heritage BiTD	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	$U = 73.5$ $p = .018$ $r = .481$
MoTD vs. IL2 BiTD	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Heritage BiTD vs. IL2 BiTD	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	$U = 29.0$ $p = .041$ $r = .426$

4.7 Discussion

The point of departure for this study was the recommendation of assessing bilingual children in both languages to avoid misdiagnosis with SLI. However, due to reduced input conditions, heritage bilinguals demonstrate high individual variation in their HL abilities, and some of their interlanguage phenomena overlap with language markers of monolingual SLI. If the HL is not well developed in both children with and without language disorder, this would limit the diagnostic potential of HL assessment. Therefore, this study set out to investigate whether HL assessment is feasible and which diagnostic tools could be employed in heritage settings. Focusing on Arabic as an HL, this study evaluated the clinical utility of a standardized test battery for Arabic (ELO-L) and an Arabic LITMUS-SRT for ruling in/out SLI in (pre)school-age bilingual children acquiring Levantine Arabic as an HL in Germany. Both tools were developed and shown to have good diagnostic accuracy in the Lebanese bilingual context, where Arabic is acquired in a majority diglossic context. However, it is still an open question whether they can reliably disentangle typical language development from SLI in bilingual children acquiring Arabic in a heritage minority context, that is, in a diglossia-less context with limited or no exposure to Standard Arabic via formal education or literacy boosting programs. To address this issue, the study compared the performance of heritage bilingual children with and without SLI to that of MoTD and late successive bilingual children as control groups on the two tests.

The study first asked how the groups compared on the subtests of the standardized test battery ELO-L and on the Arabic LITMUS-SRT, and whether the two tests could be applied with equal success to rule in/out language disorder in heritage contexts. The second objective was to examine whether the performance of the typically developing bilinguals on the tests was affected by child-internal and language environment factors known to influence the performance of bilingual children on language measures. For the Arabic SRT, the study also asked whether performance is primarily predicted by language ability or by working memory, where an interaction between working memory and language ability was to be expected given that the task features computationally complex constructions featuring nonlocal dependencies that are more taxing for working memory resources. Finally, the study asked how the groups compared on the individual structures of the Arabic LITMUS-SRT and whether all structures distinguished equally well

between typically developing and language-disordered children in heritage contexts. The results of the analyses, broken down by task, are discussed below.

4.7.1 Standardized HL test battery for Arabic

Results of the ELO-L subtests show that, despite outperforming heritage BiSLIs on all subtests, heritage BiTDs displayed different ability profiles from their typically developing monolingual and late successive bilingual counterparts. As demonstrated in Figures 6-11 and Tables 8/9, heritage BiTDs scored lower than MoTDs and late successive BiTDs on all ELO-L subtests except for comprehension of morphosyntax and phonology, where all TD groups performed alike. Specific weaknesses were observed for the subtests Receptive and Expressive Vocabulary. Although the performance of the heritage BiTDs on receptive vocabulary was slightly better than on expressive vocabulary, both measures appeared to be highly sensitive to reduced HL input effects. Vocabulary was negatively influenced by an earlier exposure to the societal language and the amount of current L2 use, which explained a significant proportion of the variance in the performance of the BiTD. A later AoO to the L2, which coincides with the onset of reduced exposure to HL input, was associated with better vocabulary scores, especially for the expressive measure, where it accounted for more than 50% of the variance. Expressive vocabulary was additionally predicted by the quality of parental input, which affects the quantity and the richness of the input heritage bilingual children receive. Parents who have higher proficiency levels in the HL and are themselves first-generation immigrants are likely to use the HL more frequently with their children and provide qualitatively enriched and differentiated HL input (Daskalaki et al., 2020).

The exceptionally low performance of heritage bilingual children on HL vocabulary aligns with findings of previous studies showing that vocabulary is very susceptible to input effects in bilinguals (Balilah, 2017; Bialystok et al., 2010; Chondrogianni & Marinis, 2011; Oller et al., 2007; Thordardottir, 2019). The sensitivity of vocabulary to reduced input effects has been ascribed to the distributed and context-specific characteristic of vocabulary (Bedore et al., 2005). Despite gains in HL vocabulary over time, bilingual children constantly face the challenge of trying to catch up to a “moving target” (Cummins, 2000) in that while bilingual children are trying to bridge the vocabulary gap, monolingual children are in the process of expanding their

vocabulary. This becomes much more pronounced after the onset of schooling in the societal language and the gradual shift in language dominance, especially when HL children lack access to literacy programs and are not given the opportunity to learn about the different “registers and the pragmatic functions of [their heritage] language in different contexts and for different purposes” via text exposure (Montrul, 2022, p. 403). Taken together, the latter findings highlight the danger of sole reliance on measures of vocabulary for the assessment of bilingual children in the HL.

The measure in which all TD groups did equally well, reception of morphosyntax, which appeared to be more robust against variation in HL exposure than vocabulary. On the other hand, production of morphosyntax proved to be sensitive to input variables, though to a lesser extent than vocabulary. Performance of the BiTD children was predicted by AoO of exposure to the L2, early exposure to the HL and L2, and current L2 use (see Albirini, 2018, for similar results). Interestingly, despite outperforming their heritage BiTD peers, late successive BiTDs did not perform at ceiling on this subtask, which is noteworthy. A subsequent inspection of the children's errors on this subtest showed that both groups particularly struggled with items eliciting nonconcatenative plural forms, which represent a late-acquired phenomenon in monolingual acquisition and require more input to stabilize (see Albirini, 2018, and Albirini & Benmamoun, 2014, for similar findings). Heritage BiTDs also struggled with items eliciting gender and plural agreement and double object constructions with resumptive pronominal clitics, which are reported to be vulnerable in heritage speakers (Albirini & Benmamoun, 2014; Montrul & Sánchez-Walker, 2013).

The Phonology subtest was predicted only by chronological age, which, in a sense, can be viewed as a proxy for the LoE to the HL. This is an expected outcome as the task includes several items featuring late-acquired phonemes in addition to phonemes attested in MSA but not in the colloquial varieties and reinforces earlier reports of the effect of the phonological distance between the spoken vernaculars and the standard on children's phonological memory (Saiegh-Haddad & Ghawi-Dakwar, 2017) and phonological awareness in Arabic (Saiegh-Haddad, 2003, 2004, 2007). Older children are more likely to be familiar with unique MSA phonemes through Sunday schooling and Quranic lessons. Interestingly, not only younger bilingual children but also younger preschool monolingual children, who have not been immersed in MSA education at the time of assessment, encountered problems with the latter phonemes. This is in line with previous

research showing that monolingual Arabic-speaking children continue to encounter difficulties with standard phonemes from preschool until Grade 6 (Schiff & Saiegh-Haddad, 2018) and even in higher grades in some children, for example, children from low-SES backgrounds (Saiegh-Haddad et al., 2020).

Concerning the question as to whether the ELO-L test battery can be used to rule in/out language disorder in the German heritage context, the results show that the test can be used to rule out language disorder in case a child shows adequate performance, but not vice versa, i.e., there is a risk of overdiagnosis with SLI. Results on the ELO-L showed that BiSLI children performed significantly poorly on all subtests, even on the receptive ones, not only compared to the typically developing groups but also relative to the norming population, as indicated by their very low z-scores. However, although this result appears to be promising and might imply that the test is sensitive to SLI in heritage contexts, this does not mean that the test does not overidentify typically developing heritage bilinguals as SLI if traditional diagnostic criteria are applied. In clinical reality, a speech-language therapist would usually compare a child's performance to the norm references provided in the test manual, and if a typically developing bilingual child performed below the pathology cutoff in two out of five domains, the child would be classified as SLI based on the HL. This speculation was born out for the ELO-L test employed in this study. When z-scores with reference to the norming sample were considered, a sizable proportion of the heritage BiTD (12/18) had at least two scores in the pathology range despite applying the criteria for adjusting monolingual norms with bilingual children, predominantly in expressive/receptive vocabulary and morphosyntax production. Late successive bilinguals, on the other hand, were all classified as BiTD. Hence, one can conclude that the test battery is more suited for the assessment of late successive bilinguals with limited exposure to the societal language, which is also an encouraging result since L2 assessment tools are lacking for this group (see Hamann et al., 2020).

4.7.2 Arabic LITMUS-SRT

The study further asked how heritage BiTD compared to MoTD, late successive BiTD, and heritage BiSLI on a LITMUS-SRT for Levantine Arabic, a test that can be administered in less than 10 minutes. The results are auspicious: Heritage BiTD

performed on par with MoTD and late successive BiTD on both SRT scoring measures, “identical repetition” and target structure. Importantly, the task appeared to be robust against reduced input effects as it was correlated with neither age nor input variables. Moreover, the only language measure explaining variance in performance on SRT “identical repetition” and SRT target structure was the Morphosyntax reception subtest of the ELO-L (see **Table 18**), on which heritage BiTD performed on par with their MoTD and late successive BiTD peers. These results are in line with the findings of Hamann et al.'s (2020) study, which explored the clinical utility of the task with older heritage and late successive BiTD. However, a difference to Hamann et al.'s results emerged in that verbal short-term memory (as indexed by FDS) also significantly contributed to explaining the variance in the performance of the BiTD, which is likely to be related to the younger age of the bilingual participants featured in this sample, whose working memory capacities are still developing. This finding aligns with previous studies on L2 LITMUS-SRTs, confirming that processing computationally more complex structures, such as those included in the Arabic LITMUS-SRT, necessitates more working memory resources, which might be even more taxed when bilingual children have to perform in their weaker language (cf. Chapter 6 of this thesis; Zebib et al., 2020b).

Group comparisons of the individual structures showed that BiSLI performed significantly worse than the TD groups on all structures for both scoring measures. Heritage BiTDs, on the other hand, showed comparable performance to their MoTD and late successive BiTD counterparts on all structures except for object relative clauses, on which they did not significantly differ from the BiSLI group either. This finding contrasts with Saad and Henry's (2020) results, who found object relatives to be among the structures that best distinguished BiSLI from BiTD in the Lebanese majority context. That object relatives are problematic for heritage BiTD aligns with Albirini's (2018) findings showing them to be challenging for heritage Arabic–English BiTD in the United States. Despite the lack of a statistically significant quantitative difference between heritage BiTD and BiSLI on object relatives, additional qualitative analyses revealed differences between the two groups. BiSLI mainly produced sentential fragments (48/54 errors), indicating their inability to cope with computational complexity (see also Taha et al., 2021b). In contrast, heritage BiTD opted for complex constructions and mainly omitted the resumptive object clitic within the relative clause, producing subject relative clauses (5/26 errors) or finite complement clauses in place of the object

relatives (8/26 errors). A further infrequent error type was producing indefinite subject relatives (3/26) or omission of the relative pronoun (4/26 errors). Despite its theoretical relevance, the limited speech-language pathologists' resources would not allow for such a fine-grained qualitative analysis, implying that it is best to avoid including object relatives in Arabic SRTs targeting heritage populations, at least for the age range investigated in this study.

An important finding is that the identical repetition scoring scheme did not disadvantage heritage BiTD. This is an encouraging result as this scoring measure is faster and easier than target structure. One unexpected result was the lack of a significant difference between late successive BiTD and heritage BiSLI on finite complement clauses; however, a significant difference emerged once the sentences were scored by correct target structure. Interestingly, only the typically developing children profited from scoring the sentences by target structure. The primary source for the typically developing children's target-conform yet nonverbatim repetitions on the task were omissions and additions of the definite article [-al]. In most cases, such errors did not affect the realization of the targeted structure; however, omitting the definite article of the relativized noun in relative clauses renders them non-target-like. Notably, most omissions of the definite article concerned words that begin with coronal consonants. When the definite article is added (as a prefix) to a word beginning with a coronal sound in Arabic, the lateral sound of the definite article is fully assimilated into the following coronal sound resulting in a geminate sound, for example, el + febe:k "the + window" → eʃ.ʃebe:k. Heritage speakers are reported to have problems with the perception of geminate sounds (see Alkhudidi et al., 2020) and might hence not realize that such words are definite.

4.7.3 HL abilities of bilingual children with SLI

As shown in the preceding sections, heritage bilingual children with SLI performed very poorly on LITMUS-Arabic SRT and on all ELO-L measures, even the receptive ones, not only compared to the investigated groups but also relative to the norming population. Comparing the performance of the same children to L2 measures shows a discrepancy between their language abilities in the HL and L2, with inferior abilities in the HL, especially on expressive language measures, where some children of the BiSLI group

were not able to complete the expressive measures or responded in German. According to Gillam et al. (2019), the reason for this asymmetry is that children with SLI encounter difficulties processing language input and need more input to reach a given language level (see also Blom et al., 2022). This reflects a frequent problem, namely, that parents of bilingual children with SLI are often advised to refrain from using the HL with them and restrict their interactions to the societal language. Hence, bilingual children with SLI do not have the same opportunities as their peers with typical development to develop their HL skills “following their own trajectory of language development” (Novogrodsky & Meir, 2022, p. 22). Existing HL boosting programs are not tailored to meet the needs of bilingual children with SLI and HL-speaking speech-language pathologists, and HL intervention materials are lacking. Consequently, heritage children with SLI are at significant risk of losing their HL (see Blom et al., 2019, 2022).

4.8 Conclusion

Overall, the results have demonstrated that evaluating HL abilities alongside the societal language could assist clinical decisions; however, caution is needed when choosing HL assessment tools. In the case of late successive bilingual children with extensive premigration experience in the HL and limited exposure to the L2, standardized assessment appears to be feasible and is likely to be reliable. However, standardized formal assessment of the HL (even with norm adjustments) disadvantages heritage bilingual children, especially in the domains of expressive/receptive vocabulary and morphosyntax production, which appear to be sensitive to variations in HL input. This implies that sole reliance on standardized tests developed for majority contexts is not suitable for the assessment of heritage bilinguals and that standardized HL measures can be, at best, used to rule out but not to rule in SLI in the case of heritage bilingual children. On the other hand, the Arabic LITMUS-SRT does not disadvantage heritage children. Hence, the task could be administered as a first estimate of HL abilities and can be used to provide evidence that heritage and late successive bilingual children have typical development in their HL. Despite the promising results for the bilingual sample investigated in the current study, it should be mentioned that a ceiling effect was observed for the task in heritage BiTD older than age 9;0 in Hamann et al.'s (2020) study. This

implies that the task might be too easy for older children and needs to be verified with older heritage BiSLI to verify whether the task is discriminatory for this age group. Moreover, task construction should consider aspects known to cause difficulties in heritage contexts and pay special attention to the morphophonological features of the words used to create the test stimuli. Finally, in order to avoid HL loss in bilingual children with SLI, educators and clinicians need to increase the parents' awareness of the importance of not discontinuing providing HL input to their bilingual children with SLI, as this does not exacerbate their language disorder.

5 Study 2: Influence of Language Dominance on LITMUS-Repetition tasks¹⁶⁰

Abstract

Given the great variability in bilingual language exposure and the potential effect of language experience on language performance in bilingual children, this study examines whether background variables related to bilingualism, particularly the degree of language dominance as measured by the relative amount of use and exposure, could compromise the diagnostic accuracy of the German LITMUS SRT and QU-NWRT. As described in 2.3.1 and 3.2.3, both tasks incorporate phonologically and syntactically complex structures shown to be cross-linguistically challenging for children with specific language impairment (SLI) and aim at minimizing bias against bilingual children while still being indicative of the presence of language impairment across language combinations. The study further investigated whether combining the two tasks provides better diagnostic accuracy and helps avoid misdiagnosis.

Twenty-one monolingual and 56 bilingual children (age 5;6–9;0 years) with and without SLI were included in the study. An unsupervised machine learning algorithm, the Partitioning-Around-Medoids (PAM, Kaufman & Rousseeuw, 2009), was used for deriving a clinical category for the children as $\pm$ language-impaired based on their performance scores on the LITMUS SRT and QU-NWRT (in isolation and combined) while withholding information about their clinical status based on standardized assessment in their first language (L1, also known as heritage language) and second language (L2, also known as societal language). Subsequently, we calculated diagnostic accuracy and used regression analysis to investigate which background variables (age of onset, length of exposure to L2, degree of language dominance, socioeconomic status (SES), and risk factors for SLI) best explained the clinical-group-membership yielded

¹⁶⁰ This chapter contains substantial (verbatim) parts from Abed Ibrahim and Fekete (2019), which was published in *Frontiers for Psychology*. Full reference: Abed Ibrahim, Lina & Fekete, István (2019). What Machine Learning can tell us about the role of language dominance in the diagnostic accuracy of German LITMUS non-word and sentence repetition tasks. *Frontiers in Psychology*, 9, 27-57. For this reason, the editorial “we” is used throughout this chapter. The author of this thesis developed the theoretical framework, gathered, evaluated the data and composed most parts of the manuscript, while István Fekete performed statistical analyses and helped compose the description of the statistical procedure concerning PAM clustering.

from the PAM-analysis based on the children's NWRT and SRT performance scores. Results show that risk factors for SLI were significant predictors for clinical group membership in all models, whereas language dominance did not contribute at all to explaining clinical cluster membership as typically developing or SLI based on any of the combinations of the SRT and NWRT variables. This confirms that the diagnostic accuracy of the tools is not compromised by language dominance despite the evident influence of language dominance on the performance of bilingual typically developing children in the SRT. The results further reveal that a combination of SRT scored by correct target structure and the structurally more complex language-dependent part of the NWRT yields better diagnostic accuracy than single measures.

5.1 Introduction

Even in simultaneous bilingual language acquisition, bilingual children are likely to receive less exposure to each of their languages than monolingual age-peers acquiring the respective languages. As a consequence, the linguistic profiles of bilingual children are often unbalanced, i.e., at a given age, their language abilities are unevenly distributed across and within language domains (Döpke, 2000; Kohnert, 2010; Yip & Matthews, 2007). The language with the more advanced state of development within the process of language acquisition (Deuchar & Muntz, 2003; Gathercole, 2016; Genesee & Nicoladis, 2007) or the language to which the child receives more input on a regular basis (Pearson et al., 1997) is commonly described as the dominant (stronger) language as opposed to the weaker or non-dominant one (see also, Meisel, 2007). In this sense, dominance is associated with language exposure/use (Grosjean, 2016) and/or with the degree of proficiency in either language (Deuchar & Muntz, 2003; Genesee & Nicoladis, 2007; Petersen, 1988). In the present study, Argyri and Sorace's (2007) definition of dominance is adopted, according to which the dominant language is "the language in which the bilingual child obtains more input on a regular basis" (p. 83); see also Grosjean (2010). Language dominance can shift over time due to changes in the quantity and quality of L1/L2 input due to structural changes in the child's immediate family, schooling, or place of residence (Paradis, 2010a, 2023). For example, in the case of early successive bilingual children who start acquiring the societal language (L2) while their home language (heritage language or L1) is still at an early developmental stage, a change in the degree

of dominance is frequently observed with schooling (cf. De Houwer & Bornstein, 2016a; Flores, 2015). Diagnostic problems particularly occur when bilingual children are solely assessed in the majority/societal language, which might still be their weaker language at the time of assessment leading to overdiagnosis with SLI (Bedore & Peña, 2008; Grimm & Schulz, 2014). To this end, this study examines whether background variables related to bilingualism, particularly the degree of language dominance as measured by the relative amount of use and exposure, could compromise the diagnostic accuracy of the German LITMUS SRT and QU-NWRT, which aim at minimizing bias against bilingual children while still being indicative of the presence of language impairment across language combinations. The study further investigates whether combining the two tasks provides better diagnostic accuracy and helps avoid misdiagnosis.

Several recent studies (e.g., Armon-Lotem & Meir, 2016; Boerma & Blom, 2017; Chilla et al., 2021; Meir & Armon-Lotem, 2016, 2017; Tuller et al., 2018) investigated the diagnostic potential and impact of different variables related to bilingualism on the performance in LITMUS-SRTs and NWRTs. Here, we report on three studies of direct relevance to the present study that were conducted within the joint German-French project (BiLaD) using a similar methodology with bilingual groups (Arabic/Portuguese/Turkish as L1) in Germany and France, who vary in their sociolinguistic settings. De Almeida et al. (2017) investigated the diagnostic accuracy of French LITMUS-SRT and QU-NWRT, which were constructed in close parallel to the German LITMUS SRT and QU-NWRT. They examined whether factors of L2 language use and exposure influenced the bilingual children's performance on the tasks. Although both tasks significantly discriminated between SLI and TD in both monolingual and bilingual children, reduced specificity of SRT was observed for children not dominant in French. Significant correlations were found between SRT performance and language use and dominance in the BiTD but not in the BiSLI group, suggesting that dominance might be responsible for the variation observed in the BiTD group. To avoid cases of overdiagnosis and enhance diagnostic accuracy, the authors recommend combining SRT with NWRT, which did not correlate with any of the L2-exposure variables. In a further study, Tuller et al. (2018) reported on direct comparisons between German and French LITMUS SRTs and QU-NWRTs. Their results showed good to excellent diagnostic accuracy in monolinguals, whereas the diagnostic accuracy for bilinguals was fair to good, i.e., the tasks generally distinguished bilingual children likely to be language-impaired from those likely to be typically developing. The authors further explored whether performance on the two tasks was

mainly ascribed to developmental risk factors for SLI or to factors related to bilingualism. Results showed that a sizable proportion of the variance in the performance of the bilinguals (BiSLI and BiTD collapsed together) in the German and French LITMUS-SRTs and NWRTs was explained by risk factors for SLI as measured by the index of *Positive_Early_Development* (see 3.2.3.3 for details). Exposure and use variables, such as current L2-richness, accounted for an additional 4% of the variance in the French-SRT and 11% of the variance in the German SRT. For the German QU-NWRT, early L2-exposure weighed negatively to account for a further 7% of the variance. Since current L2-richness and early exposure to L2 both contribute to establishing language dominance based on the PaBiQ (see 3.2.3.3), this raises the question of whether language dominance compromises the diagnostic accuracy of the LITMUS tools, especially the LITMUS-SRT. This question was further pursued in Hamann and Abed Ibrahim (2017), who used k-means cluster analysis to group bilingual children based on their performance scores on the German LITMUS-SRT and NWRT as language-impaired or not, without access to their clinical group membership based on standardized assessment. In order to measure diagnostic accuracy, the children's k-means cluster membership based on SRT and NWRT scores was compared to the likelihood of a child having SLI or TD based on standardized assessment in each of the child's languages. While the sensitivity rates for both SRT (scored by identical repetition, *SRT_Id*) and NWRT were excellent, the specificity rates were only suggestive, as several bilinguals were assigned to the clinical cluster based on their global NWRT and *SRT_Id* scores. In line with previous studies on German LITMUS-SRT, e.g., Hamann et al. (2017), Hamann and Abed Ibrahim (2017) showed that using the rating measure "target structure" (*SRT_Tar*), which assesses the preservation of the constructions targeted by the task, resulted in better specificity and better overall diagnostic accuracy than *SRT_Id* in the bilingual groups. The individual scores of the children likely to be BiTD were plotted against language dominance for each task. While LITMUS-QU-NWRT appeared to be rather unaffected by language dominance, 25% of the L1-dominant children performed below cutoff even on *SRT_Tar*. Finally, the study showed that a combination of SRT and NWRT helped avoid cases of over-identification with SLI among bilingual children.

Since language assessment of bilingual children is generally carried out in the societal language, the finding that dominance appears to impact the SRT performance of bilingual typically developing children raises the question of whether the task is suited for the identification of SLI in L1-dominant children when administered in their non-

dominant, i.e., weaker language German. It should, however, be noted that the three studies above have their limitations. In all of them, the diagnostic accuracy of the tools was measured against established clinical status based on standardized evaluation in the L1 and L2, not considering cases of selective impairment or problems with L1 standardized tests in heritage contexts, as also attested in Chapter 4 of this thesis. This, in turn, might be responsible for the reduced accuracy rates (see de Almeida et al., 2017, and Hamann & Abed Ibrahim, 2017, for a discussion). Hamann and Abed Ibrahim (2017) showed that using an alternative procedure that considers selective impairments and problems with L1-assessment in minority contexts minimized the slight overlap between BiTD and BiSLI and enhanced diagnostic accuracy. A further limitation is that in de Almeida et al. (2017) and Hamann and Abed Ibrahim (2017), dominance was not included as a variable in the regression models and may have been confounded with other factors. Hence, the presumed influence of dominance remains a conjecture that needs to be statistically validated.

5.2 The present study

In line with recent research and building on work conducted within the BiLaD project, this study investigates the diagnostic potential of LITMUS sentence and nonword repetition tasks for identifying language impairment in bilingual populations. Since both LITMUS-SRT and QU-NWRT were designed to minimize bias against bilingual populations while being indicative of the presence or absence of language impairment, the following research questions emerge in light of the previous findings:

- i. Upon sufficient exposure to the L2, how robust are German LITMUS-SRT and QU-NWRT against language dominance? Are they only sensitive to risk factors for SLI, or could background variables related to bilingualism, particularly the degree of language dominance (estimated by the relative amount of use of and exposure to L1/L2), compromise their diagnostic accuracy?
- ii. Since a combination of tools evaluating different aspects of language ability, such as morphosyntax and phonology, is recommended to acknowledge the heterogeneity within the SLI population and avoid cases of over- and

underdiagnosis, does a combination of LITMUS-SRT (especially when scored by correct target structure) and LITMUS-QU-NWRT yield higher diagnostic accuracy rates than those estimated for each of the tasks in isolation?

- iii. Does a combination of SRT_Tar, which evaluates the mastery of complex constructions, and the phonologically more complex language-dependent part of the LITMUS-QU-NWRT provide better diagnostic accuracy for identifying SLI in (monolingual) and bilingual children than other combinations of measures?

To address these questions, an unsupervised machine learning algorithm was used, the Partitioning around Medoids (PAM, Kaufman & Rousseeuw, 2009), for deriving a clinical category (clustering) for the children as $\pm$ language-impaired. The clustering is based on the participants' performance scores on German LITMUS sentence and nonword repetition tasks (in isolation and combined) while withholding information about their clinical status as established by standardized assessment in L1 and L2. Subsequently, the diagnostic accuracy of the tasks was calculated (separately and combined) by verifying the goodness of the fit against the clinical status of the bilinguals as TD or SLI, as determined by their norm-referenced L1 and L2 test scores as described in 3.1 and 3.2. Next, regression analysis was used to investigate which background variables (age, AoO_L2, LoE_L2, degree of language dominance, socioeconomic status, and risk factors for SLI) best explained clinical-group-membership based on the children's NWRT and SRT performance scores. The study's premise is that if the PAM-cluster membership can be predicted by the presence of risk factors for SLI but not by any of the other background variables known to influence the performance of bilingual children on language tests, particularly the degree of language dominance, then clustering of cases cuts across the SLI/TD dimension confirming that the German LITMUS-SRT and QU-NWRT are sensitive to language impairment and are not biased against bilingual children regardless of their language dominance.

5.3 Method

5.3.1 Establishing language dominance in bilingual children

Several methods have been put forward for measuring and operationalizing language dominance in bilingual children. These measures fall into two categories: performance-based and experiential-based (Bedore et al., 2012; Unsworth, 2015; Unsworth et al., 2018). Estimates of language dominance obtained by performance-based measures are based on quantitative differences in proficiency measurements between the two languages of a bilingual. These measures are usually extracted from a) spontaneous speech data, such as mean length of utterance (MLU), upper bound (UB, length of the longest utterance in a speech sample), multi-morphemic utterances (MMU), lexical diversity measures (number of different word types, verbs and nouns and type/token ratio) and directionality of code-mixing (see Cantone et al., 2008 and Kupisch, 2008 for overviews), and b) language proficiency measures based on standardized tests for vocabulary and grammar. Experiential measures, on the other hand, rely on biographical information and estimates of language use and exposure to predict dominance in bilingual children. The rationale behind the latter approach is that the (relative) proficiency of bilingual children in each of their languages is “in some sense a function of the amount of language to which they are exposed in these two languages” (Unsworth, 2015, p. 156). Accordingly, experiential variables like the relative amount of language use and exposure to each of a bilingual’s languages can be used to predict the degree of bilingual language dominance.

Bedore et al. (2012), Unsworth (2015), and Unsworth et al. (2018) found that the relative amount of exposure and use reliably predicted dominance group membership as determined by proficiency measures, confirming that the relative amount of use and exposure can be used as a proxy for language dominance in bilingual children. For the purposes of the present study and building upon the findings of Bedore et al. (2012), Unsworth (2015), and Unsworth et al. (2018), experiential-based measures were used to establish the participants’ language dominance employing the information obtained by the PaBiQ as outlined in **3.2.3.3**.

5.3.2 Materials and assessment tools

All monolingual and bilingual participants were assessed with a comprehensive test procedure to verify their clinical status as outlined in **3.1.1**. This included standardized tests in both the L1 and L2, considering the effect of language dominance on the performance of bilingual children in tests normed with monolingual populations, as Thordardottir (2015a) recommended. All children performed the German versions of the SRT (Hamann et al., 2013) and LITMUS QU-NWRT (Grimm et al., 2014), described in **3.2.3.1** and **3.2.3.2**, respectively. Relevant background information necessary for calculating the risk index for SLI and establishing language dominance was gathered via the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015, see **3.2.3.3** for a detailed description).

5.3.3 Participants

Except for three children, the current study uses the same participant sample as Hamann and Abed Ibrahim (2017), including 77 children, 21 German monolingual controls, and 56 bilinguals¹⁶¹ acquiring German alongside their L1s, Arabic, European Portuguese, or Turkish. The age range of the participants was 5;6-9;0 years. Thirty-five out of 56 children were simultaneous bilinguals, while 21 were sequential bilinguals whose systematic exposure to the L2 mainly started upon kindergarten entry at approximately age three. Almost all of the bilingual participants had a length of exposure to German of more than 24 months at the time of testing, with a mean LoE of 5;1 years ($SD = 1;10$). A participant overview is given in **Table 21**. In this table, the BiTD children are divided into subgroups based on their L1: Arabic = BiTD-A, Portuguese = BiTD-P, and Turkish = BiTD-T. The BiSLI group is composed of 12 children (4 with L1 Arabic, 3 with L1 Portuguese, and 5 with L1 Turkish). Due to the relatively small sample size of the latter group, bilingual children with SLI are grouped together regardless of their home

¹⁶¹ Note that by applying the modified classification criteria outlined in Chapter 3, the clinical status of 4 children who were initially classified as BiTD in Hamann and Abed Ibrahim (2017) changed to BiSLI. Two of them were cases of selective impairment in the L2, i.e., they performed below cutoff in the L1, and only showed deficits in morphosyntax in the L2 (grammatical SLI) plus SLI markers in the speech samples in both languages. The other two were L2-dominant and performed below-cutoff in all of the domains in the L2, but slightly above cutoff in the L1. Since SLI markers were present in their speech samples, they were classified as BiSLI.

language. The bilingual children are further classified according to language dominance as measured by the PaBiQ (see 3.2.3.3). As shown in **Table 21**, almost half of the children in the BiTD group (21/44) are dominant in their L1. The majority of the BiSLI children (9/12), on the other hand, are either balanced or L2-dominant. This reflects the advice frequently given to parents of bilingual children with atypical language development, according to which parent-child interactions are to be restricted to the societal language to avoid aggravating existing language difficulties, which entails less exposure to the L1.

The four groups (MoTD, MoSLI, BiTD, and BiSLI) were comparable regarding non-language variables such as chronological age and nonverbal intelligence. Concerning age, the overall effect of Group was not significant, as revealed by the Kruskal–Wallis test [$\chi^2(3, N = 77) = 5.505, p = .138, \eta^2 = .034$]. This also holds when the BiTD group is split into three subgroups by L1 [$\chi^2(5, N = 77) = 6.758, p = .239, \eta^2 = .051$]. In terms of non-verbal intelligence as measured by the CPM test, the overall effect of Group was significant [$\chi^2(3, N = 77) = 8.448, p = .038, \eta^2 = .075$]. However, subsequent pairwise comparisons using Mann–Whitney U tests controlling for false positives, that is, Type I error, revealed only one marginally significant comparison, namely MoSLI vs. MoTD ($U = 19.00, p = .06, r = .553$, Bonferroni-corrected). Nevertheless, all children in the MoSLI group had normal non-verbal intelligence. We further checked whether the bilingual groups were comparable concerning SES, AoO, LoE, and degree of L2-dominance (LDI). No significant differences emerged between BiTD and BiSLI concerning SES [$\chi^2(1, N = 56) = 2.228, p = .135, \eta^2 = .041$], AoO [$\chi^2(1, N = 56) = 3.261, p = .071, \eta^2 = .059$], LoE [$\chi^2(1, N = 56) = 0.615, p = .433, \eta^2 = .011$] and LDI [$\chi^2(1, N = 56) = 1.912, p = .167, \eta^2 = .035$]. This also holds when the BiTDs are split by L1: SES [$\chi^2(3, N = 56) = 3.216, p = .360, \eta^2 = .06$], LoE [$\chi^2(3, N = 56) = 3.640, p = .303, \eta^2 = .07$] and LDI [$\chi^2(3, N = 56) = 4.457, p = .216, \eta^2 = .08$]. Concerning AoO, the overall effect of Group was significant when BiTDs were divided by L1 into three subgroups L1 [$\chi^2(3, N = 56) = 11.833, p = .008, \eta^2 = .17$]. Mann–Whitney U tests applying Bonferroni-adjustment of p -values revealed significant differences in AoO between BiTD-A and BiTD-P ($U = 33.00, p < .05, r = .531$) as well as between BiTD-A and BiSLI ($U = 17.00, p < .05, r = .617$). Nevertheless, the overall effect of Group was not significant when the BiTD groups were collapsed together [$\chi^2(1, N = 56) = 3.261, p = .071, \eta^2 = .059$].

Table 21: Participants, including monolingual controls, after verification of clinical status (Mean, SD, and range)

Simult./total	MoTD (n=10)	MoSLI (n=11)	BiTD (n=44)			BiSLI (n=12)	
			BiTD-A (n=10)	BiTD-P (n=18)	BiTD-T (n=16)	Total (n=44) 24/44	11/12
Age at testing (in months)	75.90 (8.99) 66-92	79.00 (9.89) 68-98	88.60 (13.5) 70-108	83.44 (14.72) 66-108	86.68 (13.01) 70-104	85.6 (13.60) 66-108	80.66 (15.05) 64-108
Age of Onset of exposure to L2 (AoO, in months)	0	0	39 (13.61) 24-75	17.61 (24.39) 0-90	25.31 (15.52) 0-48	25.27 (20.65) 0-90	16.00 (17.03) 0-36
Length of Exposure to L2 (LoE, in months)	75.90 (8.99) 66-92	79.00 (9.89) 68-98	53.2 (19.68) 32-97	67.97 (24.59) 18-101	61.37 (19.01) 34-96	61.73 (21.87) 18-101	62.00 (22.77) 30-88
Non-verbal IQ: CPM (percentile rank)	81.20 (13.98) 56-100	53.72 (24.50) 25-99	44 (29.13) 9-93	73.77 (17.69) 42-94	71.18 (22.23) 38-100	66.06 (25.01) 9-100	55.08 (26.69) 27-98
LDI^a (/50)	N/A	N/A	-7 (9.34) -25-11	1.44 (13.76) -25-23	-5.75 (12.57) -27-18	-3.09 (12.77) -27-23	0.58 (12.37) -21-24
L1-dominant (no./total)	N/A	N/A	(5/10)	(8/18)	(8/16)	(21/44)	(3/12)
Balanced (no./total)	N/A	N/A	(1/10)	(3/18)	(4/16)	(8/44)	(5/12)
L2-dominant (no./total)	N/A	N/A	(4/10)	(7/18)	(4/16)	(12/44)	(4/12)
Yrs. educ. Mother (SES^b)	N/A	N/A	14.2 (1.68) 12-18	13 (3.91) 4-18	13.68 (5.23) 5-20	13.52 (4.05) 4-20	11.83 (2.75) 8-16

Notes: ^aLDI: Language Dominance Index; ^bSES: socioeconomic status; MoTD: monolingual typically developing; MoSLI: monolingual with language impairment; BiTD: bilingual typically developing; BiSLI: bilingual with language impairment.

5.3.4 Data analyses and statistical treatment

The children's responses on the SRT and NWRT were recorded using special dictaphones. Data transcription, verification, and scoring were done offline (as outlined in 3.2.3.1 and 3.2.3.2) by two independent linguistically trained raters (percentage of agreement was at least 90%). For each repetition measure, the percentage of correct responses was used as a basis for data analysis. Null reactions were counted as errors unless they were due to technical problems or errors by the investigators (missing data, less than 1% of the overall data).

Statistical analyses were conducted using IBM SPSS 24 (2016) and R-Studio (2012). Non-parametric tests were used for group comparisons due to unequal sample sizes and violation of the normality assumption, checked by the Shapiro-Wilk test. Since the study primarily aims at investigating whether the LITMUS repetition tools are suitable for the assessment of bilingual children in their weaker language, we first checked for group differences between L1-dominant BiTDs and their monolingual, balanced, and L2-dominant TD peers and whether the performance of L1-dominant BiTDs overlapped with that of MoSLIs and BiSLIs. Here, the BiTDs were split into three subgroups based on their language dominance index (LDI) as established in 3.2.3.3, and Kruskal-Wallis tests and Mann-Whitney U tests with Bonferroni adjustment were run. Recall that the BiSLIs were collapsed into a single group due to the small sample size. Since the performance of the BiTDs on SRT appeared to be influenced by dominance, partial correlation analyses controlling for age on their SRT_Id and SRT_Tar scores were conducted. In addition to language dominance, we also checked for correlations with AoO, LoE, and socioeconomic status (SES) as measured by years of maternal education since the latter factors are known to influence performance on linguistic tasks. Next, linear regression models for predicting the performance of the BiTDs on SRT_Id and SRT_Tar were built using the variables that yielded significant correlations.

Secondly, cluster analyses were applied to the data to automatically group the children into $\pm$ language-impaired based on their performance scores on the SRT (SRT_Id, SRT_Tar) and NWRT (NWRT_global, NWRT_LI, NWRT_LD), separately and then in combination. A clustering algorithm classifies a dataset into several meaningful homogenous sub-categories - so-called clusters (i.e., TD vs. SLI in this study) - based on the values of their attributes (i.e., linguistic variables in the present study) such that the similarity among objects within a category is larger than that between categories. We

opted for unsupervised learning (cluster analysis) for verifying diagnostic accuracy and establishing cutoff points separating TD from SLI on the tasks since it does not use predefined clinical status during the statistical analyses and is thus unbiased by any given classification of participants.

Because children were measured based on performance scores on LITMUS-SRT and NWRT designed to identify SLI without penalizing bilinguals, our premise was that SLI-cases would be similar to each other and hence group together, while TD-cases would form their own cluster regardless of bilingualism. Different from Hamann and Abed Ibrahim (2017), we chose the PAM (Partitioning Around Medoids) non-hierarchical k-medoid clustering method (Kaufman & Rousseeuw, 1987, 2009) over k-means because it is a suitable method for small datasets with up to approximately 60 objects and because it can handle noisy data and outliers (Kaufmann & Rousseeuw, 1987, 2009; Kashef & Kamel, 2008; Patel & Singh, 2013; Soni & Patel, 2017). Variables were scaled for normalization purposes in the course of the PAM analyses. We used the function *pam* of the *cluster* R package (Maechler et al., 2017).

Hopkins statistic (H) based on the *factoextra* R package (Kassambara & Mundt, 2017) was employed to assess clusterability, i.e., as a measure of cluster tendency (Hopkins & Skellam, 1954). If the H-value is close to zero and far below 0.5, then the dataset is clusterable (Kassambara & Mundt, 2017; Krishna et al., 2018). Because H is run on the created random dataset every time, fluctuations in the H-values might emerge if the statistics are run multiple times. Banerjee and Davé (2004) demonstrate that random data sets, clustered data sets, and regularly spaced data sets show H-values of around 0.5, 0.7-0.99, and 0.01-0.3, respectively. Because the k-medoid algorithm requires the number of clusters to be pre-defined, we first ran the *Gap Statistic* (Tibshirani et al., 2001) to determine the optimal number of clusters. The *Gap Statistic* compares the change in within-cluster dispersion for each clustering solution (at each number of clusters) to that expected at random distribution. We used the functions *fviz_nbclust* of the *Factoextra* R package (Kassambara & Mundt, 2017) and *NbClust* of the *NbClust* R package (Charrad et al., 2014) to determine the optimal number of clusters.

The k-medoid algorithm selects one of the cluster members as the most representative object, named the *cluster medoid*, so that each cluster has only one medoid. By choosing an actual case (i.e., an SLI or a TD child) as the cluster medoid, the k-medoid method is less sensitive to outliers, as mentioned before. The optimal cluster is achieved by minimizing the sum of squared Euclidean distances to the medoid in each cluster, also

called the error sum of squares (Kaufman & Rousseeuw, 1987). First, in the so-called “Build-step”, the k-medoid algorithm selects k medoids randomly, with k being the optimal number of clusters. Next, a matrix of dissimilarity is calculated from the raw data, and the algorithm assigns every object to either of the k clusters based on its distance to the nearest medoid (Patel & Singh, 2013). The sum of absolute error in the clustering procedure equals the sum of the distances between data points and their medoids. In the so-called “Swap-step”, each non-medoid object is tested as a potential medoid within each cluster by checking if the sum of within-cluster distances gets smaller if that object is used as the new medoid. If this is the case, then that configuration is used. The algorithm checks whether the solution is better than the previous one at each iteration step. If the medoids do not change, the algorithm terminates (see Patel & Sing, 2013, for details).

Because the medoid of each cluster can be seen as a prototype of that cluster, identifying the medoid can serve as a cue to interpret the cluster. For example, if the medoid of a cluster was originally diagnosed as an SLI case, then that cluster most probably represents the SLI cases. We expected the SLI-cluster to contain the majority of the children classified originally as SLI based on standardized assessment, while the majority of TD cases would reside in the other larger cluster. Our further premise was that the cluster with the lower scores on the linguistic variables would represent the cluster with language impairment since language-impaired cases score lower on the linguistic variables.

After clustering the sample, we determined the estimated cutoffs on the linguistic variables (i.e., SRT and NWRT) between the SLI- and TD-clusters based on the clustering result. A cutoff is a value of a variable that can be seen as the best threshold score to separate the cases belonging to the two categories using that variable. If the two categories can be best separated along multiple variables simultaneously, e.g., SRT and NWRT combined, then cases can be predicted (as TD vs. SLI in our study) based on multiple cutoffs on these variables. To this end, we employed conditional inference tree models (Tagliamonte & Baayen, 2012). Conditional inference trees (ctrees) are non-parametric regression models visualized as decision trees. They are suitable for our dataset because of the presence of high-order interactions among the variables and the overall small sample size compared to the number of predictors (Levshina, 2015). Besides determining the cutoff for the linguistic variables, ctrees can also provide information about the hierarchical structure of the relevant predictors of cluster membership, i.e., about variable importance. For instance, if clustering is based on several linguistic variables such as

SRT_Id, SRT_Tar, and NWRT_global, decision trees can show which one contributed the most towards predicting cluster membership as TD or SLI. The higher the variable in the hierarchy, the more important it is, with the highest-level variable being the most important. If there are multiple variables in the ctree, then a multi-hierarchy predicts the outcome (i.e., cluster membership as TD or SLI). Ctrees were implemented with the *party* R package using the ctree function (see Hothorn et al., 2006, for details).

In order to address research questions (ii) and (iii), we calculated diagnostic accuracy¹⁶² for the SRT and NWRT measures separately and combined. Sensitivity and specificity levels were estimated by comparing cluster membership of each of the children as TD vs. SLI as assigned by PAM based on LITMUS-SRT and NWRT results to their clinical status (as established by the standardized assessment procedure described in 2.4). Sensitivity is determined by the proportion of children with language impairment identified as such by LITMUS SRT and NWRT or subtests thereof (i.e., assigned to the clinical cluster in our case), while specificity is computed based on the proportion of children with typical language development identified as such by our tests, i.e., assigned to the non-clinical cluster (Dollaghan & Horner, 2011; Oetting et al., 2008). In addition, likelihood ratios¹⁶³ (LRs) were calculated based on the obtained sensitivity and specificity levels. An advantage of LRs is that they are less likely to be affected by variations in the properties of the test sample (Dollaghan & Horner, 2011). LR+, positive likelihood ratio (sensitivity/(1-specificity)), indicates how likely it is that a score below a cutoff criterion to be present in language-impaired children, while an LR-, negative likelihood ratio ((1-sensitivity)/specificity), is indicative of the likelihood of a score above a cutoff criterion to belong to a child without language impairment.

To answer research question (i), we investigated which background information variables provided by the PaBiQ as cogent confounders predicted cluster membership following each clustering procedure based on SRT and NWRT measures or combinations thereof. The hypothesis to be tested was that cluster membership as TD or SLI, based on performance scores in LITMUS-NWRT and SRT, in isolation or combined, can only be explained by variables concerning risk factors to SLI and not by background information

¹⁶² Following Plante and Vance (1994), good diagnostic accuracy is given when specificity and/or sensitivity rates are $\geq 90\%$. Rates between 80% and 90% are viewed as fair.

¹⁶³ LR+ values ≥ 10 are highly indicative of the presence of language impairment, LR- values ≤ 0.10 highly indicate the absence of an impairment, LR+ values ≥ 3.0 and LR- values ≤ 0.3 are considered to be clinically suggestive, while LR+ values < 3.0 and LR- values > 0.3 are viewed as clinically uninformative (cf. Dollaghan & Horner, 2011).

variables related to bilingualism, particularly the degree of language dominance. If this hypothesis is confirmed, then the clustering of the cases cuts across the SLI/TD dimension rather than any of the background information variables unrelated to risk factors for SLI, validating that the diagnostic accuracy of the tasks is not compromised by language dominance. To that end, Firth's Bias-Reduced Binary Logistic Regression was run (Firth, 1993), which uses penalized ML¹⁶⁴. Cluster membership (TD or SLI) served as the dependent measure. Models with Firth's correction were built using the *Brglm2* R package (Kosmidis, 2018). Given the overall small sample size, we included only a maximum of four background information variables as fixed factors in the model to avoid over-parametrization. Because regression analysis provides a way of adjusting for potentially confounding covariates included in the model, we entered the covariates into the model at once. To examine whether the diagnostic accuracy of the tasks is not compromised by language dominance and is only sensitive to risk factors for SLI, we built several regression models using Firth's correction with PAM cluster membership as TD or SLI as the dependent variable. In each model, we entered LDI and the index of Positive_Early_Development (risk factors for SLI) in addition to two other background information variables reported to explain performance on LITMUS-SRT and NWRT in other studies (see Tuller et al., 2018) as covariates. The latter variables included AoO, LoE, and SES. We also included chronological age as a covariate since working memory and cognitive capacities are rapidly growing in children and since children's language abilities tend to improve over time.

Background information variables were first scaled by the mean of their original variable to remove potential *non-essential* multicollinearity between them (Dalal & Zickar, 2011) and to adjust the interpretation of the coefficients. Multi-collinearity among covariates was checked using the *Variance Inflation Factor (VIF)* after scaling, with a VIF value above 10 indicating serious multi-collinearity (Kutner et al., 2004). Correlations between the background variables are given in the appendix.

¹⁶⁴ We used Firth's Bias-Reduced Logistic Regression to step around the following statistical concerns. Given our small sample, Long (1997) advises against maximum likelihood estimation in logistic regression with less than 100 cases. A second confound is that in small samples maximum likelihood estimates in binary logistic regression models are not powerful because there can be a complete- or quasi-complete separation along one covariate (Rainey, 2016). That is, for a given combination of covariates the outcome can be predicted perfectly. To avoid separation, Peduzzi, et al. (1996) suggest that the number of positive outcome events (i.e., the smaller number of binary outcomes) divided by the number of independent variables should be more than 10.

5.4 Results

5.4.1 Overall results on the German LITMUS-QU-NWRT and SRT: Language dominance as a factor influencing performance on LITMUS repetition tasks

Kruskal-Wallis tests comparing performance scores of L1-dominant BiTDs to the other groups (MoTD, balanced-BiTD, L2-dominant-BiTD, MoSLI, and BiSLI) on NWRT_global, NWRT_LI, NWRT_LD, SRT_Id, and SRT_Tar yielded significant results for all measures as shown in **Table 22**. Subsequently, pairwise comparisons were carried out using Mann-Whitney U tests with Bonferroni adjustment. Typically developing children performed significantly better than their language-impaired counterparts on all measures. All measures distinguished between MoTDs and MoSLIs as well as between BiTDs and BiSLIs regardless of language dominance. Moreover, all of the BiTD groups significantly outperformed MoSLIs. The comparisons yielded no significant differences between MoSLIs and BiSLIs on any of the abovementioned measures. Comparing MoTDs to the BiTDs split by dominance revealed no significant differences between MoTDs and balanced as well as L2-dominant BiTDs on either measure. Nevertheless, significant differences with large effect sizes were found between MoTDs and L1-dominant BiTDs as well as between L1-dominant and L2-dominant BiTDs for both SRT_Id and SRT_Tar but not for any of the NWRT measures (see **Table 22**). It should, however, be stressed that despite the observed significant differences in SRT_Id and SRT_Tar, L1-dominant BiTDs performed significantly better than MoSLIs and BiSLIs on both SRT measures. **Figure 15** and **Figure 16** depict the overall performance of the groups in LITMUS-QU-NWRT and LITMUS-SRT, respectively. An overview of the significant pairwise comparisons is provided in **Table 22**.

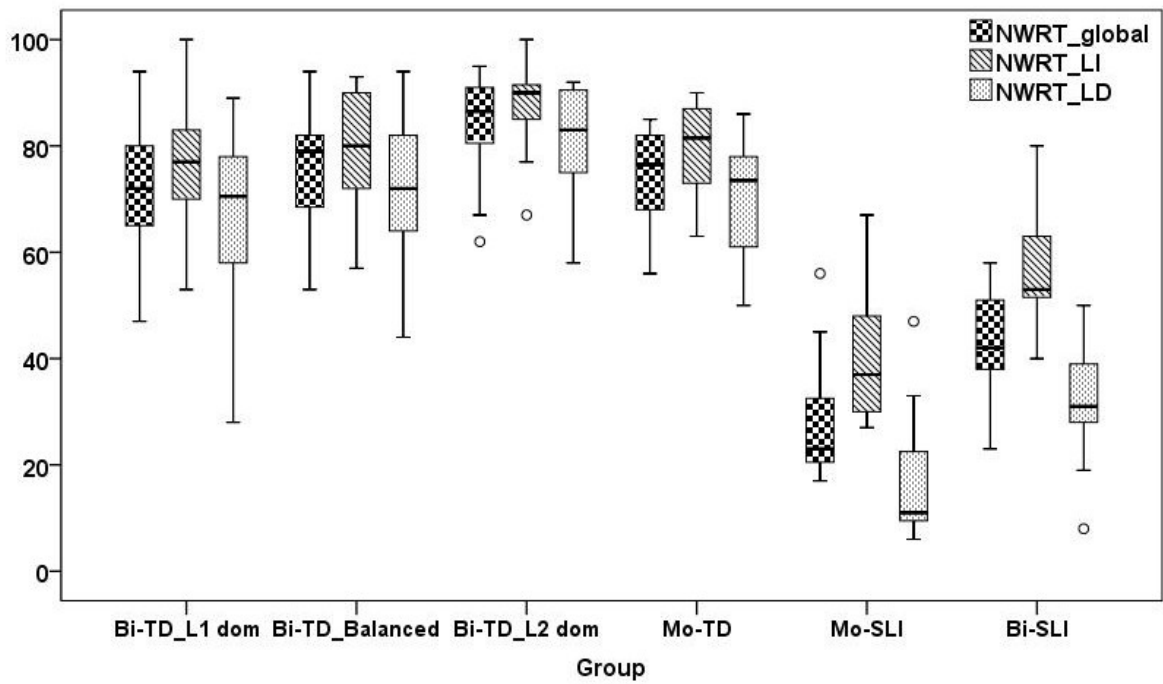


Figure 15: LITMUS-QU-NWRT: percentage of identical repetition: NWRT_global (overall score), NWRT_LI (language-independent part), and NWRT_LD (language-dependent part).

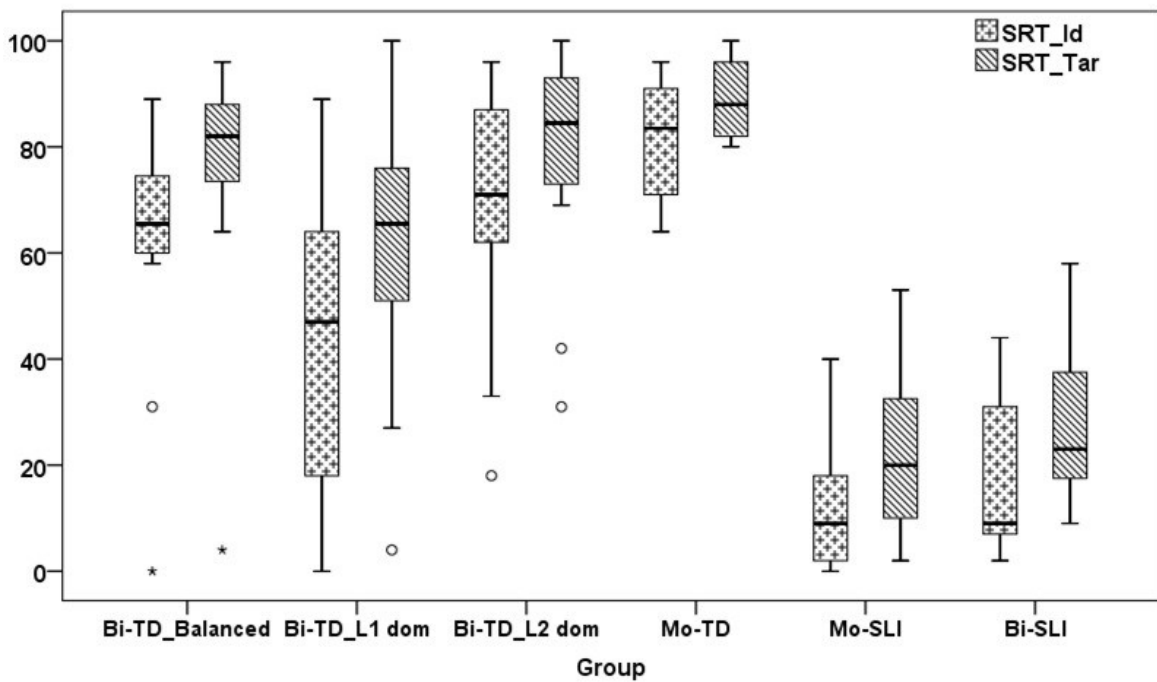


Figure 16: SRT: percentage of identical repetition (SRT_Id) and correct production of target structure (SRT_Tar).

Table 22: The effect of (clinical) group membership on LITMUS-QU-NWRT and SRT (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests)

	NWRT_global	NWRT_LI	NWRT_LD	SRT_Id	SRT_Tar
	$\chi^2(5, N = 77)$ = 39.600 $p < .001$	$\chi^2(5, N = 77)$ = 35.044 $p < .001$	$\chi^2(5, N = 77)$ = 38.714 $p < .001$	$\chi^2(5, N = 77)$ = 43.365 $p < .001$	$\chi^2(5, N = 77)$ = 43.086 $p < .001$
MoTD vs. MoSLI	$U = 0.500$ $p < .001$ $r = .838$	$U = 1.500$ $p < .001$ $r = .824$	$U = 0.000$ $p < .001$ $r = .847$	$U = 0.000$ $p < .001$ $r = .847$	$U = 0.500$ $p < .001$ $r = .846$
BiTD_L1 dom. vs. MoSLI	$U = 2.000$ $p < .001$ $r = .804$	$U = 6.500$ $p < .001$ $r = .775$	$U = 3.000$ $p < .001$ $r = .797$	$U = 33.500$ $p < .001$ $r = .592$	$U = 24.500$ $p < .001$ $r = .652$
BiTD_Balanced vs. MoSLI	$U = 7.000$ $p < .001$ $r = .775$	$U = 8.5000$ $p < .001$ $r = .763$	$U = 10.000$ $p < .001$ $r = .736$	$U = 11.500$ $p < .001$ $r = .717$	$U = 10.000$ $p < .001$ $r = .735$
BiTD_L2 dom. vs. MoSLI	$U = 3.000$ $p < .001$ $r = .845$	$U = 4.500$ $p < .001$ $r = .830$	$U = 3.500$ $p < .001$ $r = .841$	$U = 5.000$ $p < .001$ $r = .824$	$U = 5.000$ $p < .001$ $r = .822$
BiTD_L1_dom. vs. BiSLI	$U = 6.000$ $p < .001$ $r = .765$	$U = 25.000$ $p < .001$ $r = .643$	$U = 8.000$ $p < .001$ $r = .752$	$U = 44.000$ $p < .05$ $r = .512$	$U = 29.500$ $p < .001$ $r = .608$
BiTD_balanced. vs. BiSLI	$U = 2.000$ $p < .001$ $r = .802$	$U = 10.500$ $p < .001$ $r = .688$	$U = 2.000$ $p < .001$ $r = .804$	$U = 3.000$ $p < .001$ $r = .798$	$U = 0.000$ $p < .001$ $r = .829$
BiTD_L2_dom. vs. BiSLI	$U = 9.500$ $p < .05$ $r = .648$	$U = 12.000$ $p < .05$ $r = .616$	$U = 7.000$ $p < .001$ $r = .686$	$U = 4.000$ $p < .001$ $r = .726$	$U = 4.500$ $p < .001$ $r = .718$
BiTD_L1_dom. vs. MoTD	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	$U = 16.500$ $p < .001$ $r = .672$	$U = 21.000$ $p < .001$ $r = .638$
BiTD_L1_dom. vs. BiTD_L2_dom.	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	$U = 44.500$ $p < .05$ $r = .559$	$U = 53.000$ $p < .05$, $r = .501$

In the next step, all of the BiTDs were collapsed into one group, and partial correlation analysis was carried out, controlling for age on performance in SRT_Id and SRT_Tar and variables shown to influence performance on LITMUS-SRT, including language dominance (see Tuller et al., 2018). Moderate positive correlations were found between LDI and performance on SRT_Id ($r = .542, p < .001$) and SRT_Tar ($r = .586, p < .001$), as well as SES and SRT_Id ($r = .478, p = .001$) and SRT_Tar ($r = .431, p = .004$). The analysis revealed a weak positive correlation between SRT_Id and LoE ($r = .364, p < .05$) and a weak negative correlation between SRT_Id and AoO ($r = -.348, p < .05$), whereas the latter two correlations were not significant in the case of SRT_Tar. Two multiple linear regression models were built to predict the performance of the BiTDs on SRT_Id and SRT_Tar. The following variables were entered into the model as independent variables: AoO, LoE, LDI, and SES. The results show that performance on SRT_Id in the BiTD group is predicted by LDI ($\beta = 3.724, T = 2.922, p = .001$), followed by LoE ($\beta = 3.846, T = 2.287, p = .01$), and SES ($\beta = 3.424, T = 2.829, p = .001$). However, for SRT_Tar, only LDI and SES had significant effects in the full model: LDI ($\beta = 4.480, T = 3.360, p = .001$), SES ($\beta = 2.914, T = 2.301, p = .01$). The independent variables did not show multi-collinearity in the models ($VIF < 3$ for all independent variables).

5.4.2 Results of PAM-cluster analyses

Comparison of the global performance of L1-dominant BiTDs to their monolingual, balanced, and L2-dominant peers, as well as results of regression analyses show that language dominance was the first predictor to explain the performance of the BiTDs on both SRT_Id and SRT_Tar and point to the possibility that language dominance could compromise the diagnostic accuracy of the SRT if administered to bilinguals in their non-dominant language, here German. In order to examine this, we ran the k-medoid PAM-clustering to group the children into SLI vs. TD based on their performance scores on SRT and NWRT, determined the cutoff points between the clusters for each of the repetition measures, and calculated the diagnostic accuracy for different combinations of sub-measures of the two. Next, regression analyses using Firth's correction were carried out to examine whether language dominance contributed to the results of PAM-clustering, i.e., assigning the children to the clinical vs. non-clinical cluster based on performance scores on the LITMUS repetition tasks. We examined this for SRT and NWRT separately as well as combined. LDI and Positive_Early_Development were entered as predictors

for PAM cluster membership into all regression models in addition to two further background variables (age, AoO, LoE, SES).

Before applying the PAM clustering to our bilinguals, we tested it on our monolingual dataset. The following variables were simultaneously entered in the cluster analysis: NWRT_global, NWRT_LI, NWRT_LD, SRT_Id, and SRT_Tar. The *Hopkins statistic* yielded a value of around 0.23, indicating clusterable non-random data, and the Gap statistic revealed that the optimal cluster solution is 2. The clustering procedure resulted in a clear separation into two homogenous groups with two cluster medoids. The cutoff points (see 5.3.4) separating the monolingual clinical cluster from the non-clinical one in our data sample were as follows: SRT_Id: 40%, SRT_Tar: 53.3 %, and NWRT_global: 45.45%, NWRT_LI: 60%, NWRT_LD: 47.22%. **Figure 17** gives a visual representation of the k-medoid PAM-cluster analysis on monolingual data using the two-cluster solution. Cases belonging to the cluster on the right are identified as TD cases, while those in the cluster on the left are identified as SLI cases. To facilitate computing the sensitivity and specificity of the task, case numbers were combined with the clinical status as assigned by the PAM-clustering procedure. As can be seen in **Figure 17**, all of the monolingual children assigned to the MoSLI group based on standardized test procedures belong to the clinical cluster, yielding a sensitivity of 100%, whereas all of the monolingual subjects classified as MoTD based on standardized test procedures belonged to the non-clinical cluster, yielding a specificity of 100%. We also ran the clustering procedure on SRT and NWRT separately, i.e., (SRT_Id+SRT_Tar) and (NWRT_global + NWRT_LI+NWRT_LD), respectively, and obtained similar results. In the next step, we used regression analysis entering age¹⁶⁵ as a single variable to check whether chronological age could explain cluster membership. The latter analysis results indicate no association between age and the cluster variable (Firth: $\beta = -0.03167$, $Z = -0.646$, $p = .519$).

In order to check for overlap between BiTD and MoSLI, we applied the PAM-analysis to the MoSLI and BiTD children collapsed together using performance scores on both SRT and NWRT. The data yielded an H-value of around 0.18, which indicates clusterable non-random data with 2 as the optimal number of clusters. Before entering all five variables into the clustering procedure, we first performed the clustering procedure based on performance on SRT_Id + SRT_Tar. Ctree models showed that SRT_Tar, but

¹⁶⁵ SES information is only available for the bilingual participants.

not SRT_Id, predicted cluster membership with a threshold of 53.3% separating the two clusters. All MoSLI children scored below the cutoff and were thus assigned to the clinical cluster by the PAM algorithm, i.e., sensitivity = 100% with an LR+ = 6.29, whereas 37/44 BiTD children performed above threshold (specificity = 84.1%, LR- = 0.00). Age as a single variable in the regression model did not prove to be a predictor for cluster membership (Firth: $\beta = -0.02849$, $Z = -1.238$, $p = .216$). In the next step, we ran the PAM analysis on NWRT_global, NWRT_LI, and NWRT_LD. The clustering resulted in two clusters separated by a cutoff of 33.33% on NWRT_LD, which is also the primary predictor of clustering membership. 10/11 MoSLI children were assigned to the clinical cluster, yielding a sensitivity rate of 91%, and 43/44 BiTD children performed above the cutoff and were assigned to the non-clinical cluster, giving a specificity of 98% with an LR+ of 39.56 and an LR- of 0.092. Again, chronological age was not a significant predictor for the cluster variable (Firth: $\beta = -0.03618$, $Z = -1.293$, $p = .196$).

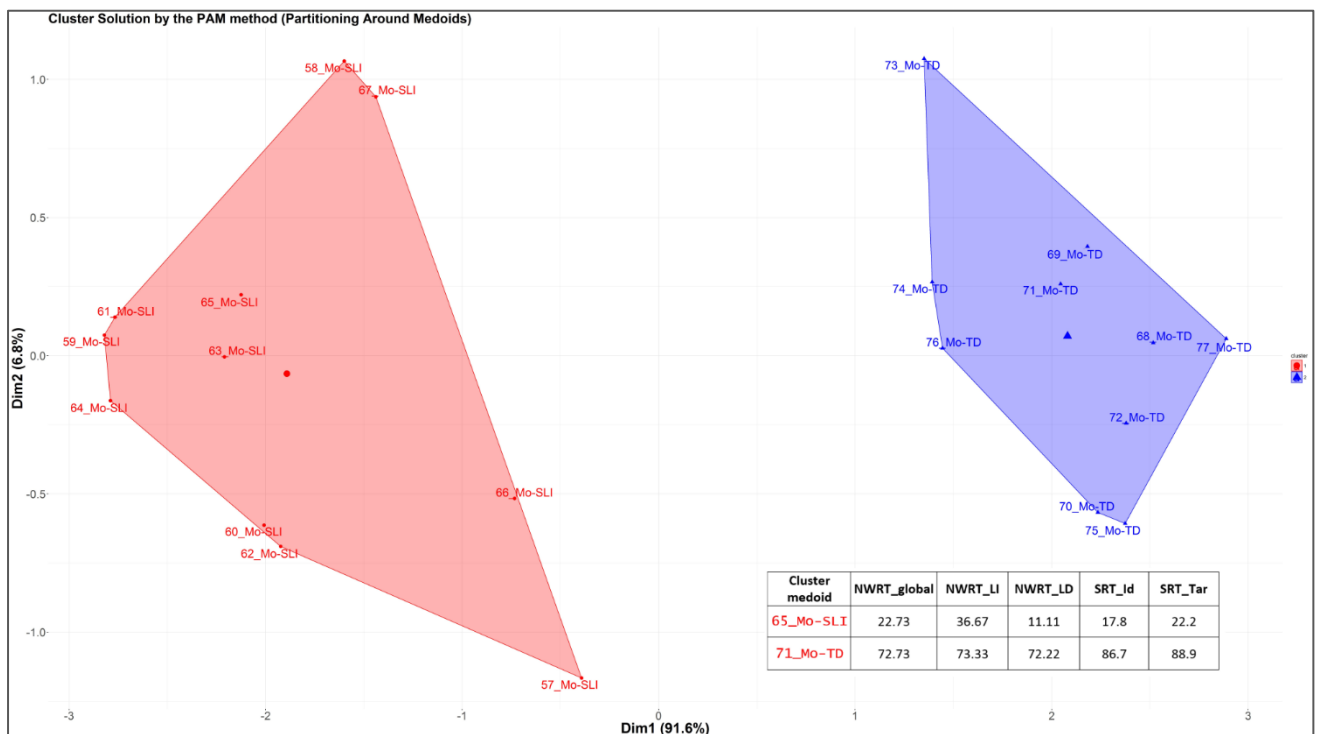


Figure 17: PAM-clustering of MoSLI+MoTD based on performance scores on SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD.

Finally, both LITMUS tasks were included in the PAM analysis using the measures NWRT_global, NWRT_LI, NWRT_LD, SRT_Id, and SRT_Tar. After entering all SRT and NWRT measures at once into the clustering procedure, 5 of the 7 BiTDs, who were assigned to the clinical cluster based on scores on SRT alone, changed membership from the clinical to the non-clinical cluster. A combination of SRT and NWRT measures yielded 100% sensitivity (all MoSLIs belong to the clinical cluster) and 95% specificity (42/44 BiTDs belong to the non-clinical cluster) with an LR+ of 20 and an LR- of 0.00. An illustration of the result of the PAM cluster analysis on the MoSLI and BiTD data is given in **Figure 18**. Age at testing as a single variable did not significantly predict PAM cluster membership (Firth: $\beta = -0.05145$, $Z = -1.824$, $p = .068$).

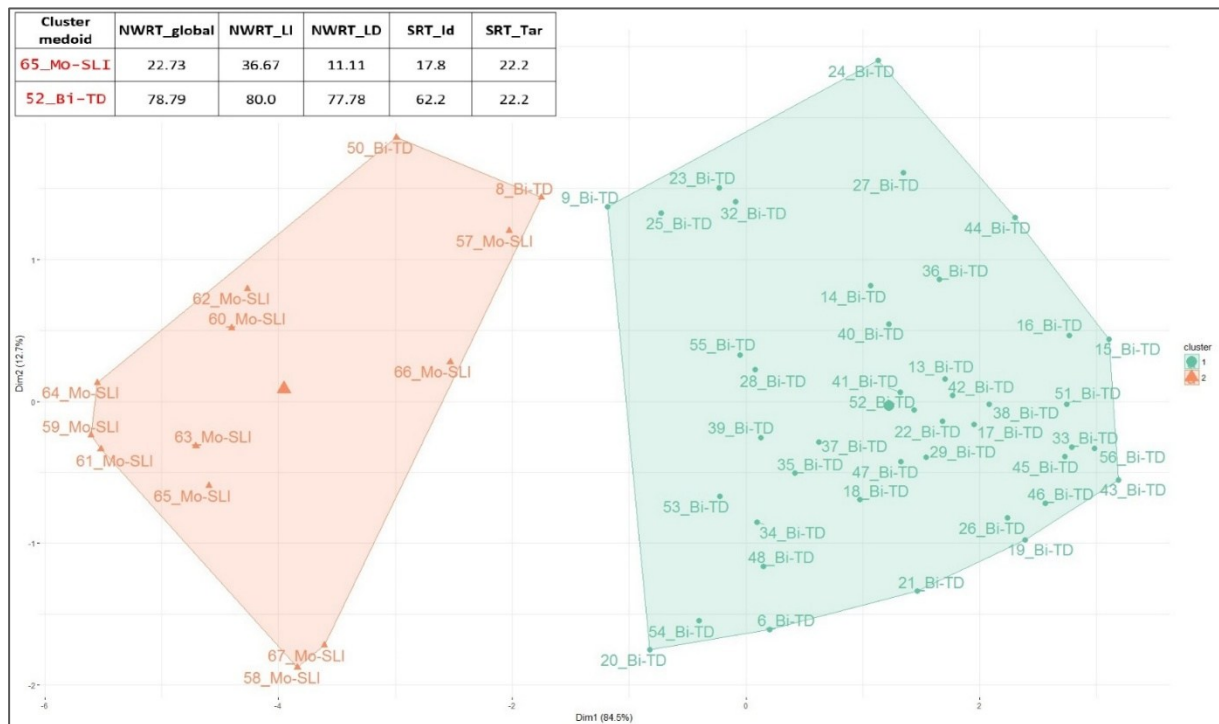


Figure 18: PAM-clustering of MoSLI+BiTD based on performance scores on SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD.

A visual representation, a ctree of the hierarchical structure of the most relevant linguistic variables for predicting PAM cluster membership, illustrated in **Figure 18**, is provided in **Figure 19**. Within a ctree, only those variables relevant to explaining the clustering results appear in the graph, where each relevant variable is represented by an oval circle, and classification rules are represented by thresholds. Classification of cases starts at the top node (root). The second most important variable is one level below the top node. Classification then proceeds by moving down the branch until one arrives at a terminal node representing classification accuracy according to PAM clustering¹⁶⁶, where classification accuracy is represented in squares (y). The two numbers next to “y” show the proportion of cases successfully classified and misclassified as SLI. The number of cases on that route is represented by “n”. Each classification route can be expressed in the form of if-then conditions with cutoffs. As can be seen in **Figure 19**, when all five measures are included in the clustering procedure, both SRT_Tar and NWRT_global are identified as significant contributors to predicting PAM cluster membership. Classification of cases starts at the top node occupied by SRT_Tar, followed by the second most important variable, “NWRT_global”, one level below the top node. Based on the hierarchical variable structure in the ctree depicted in **Figure 19**, it becomes visible that ten children whose scores on SRT_Tar were $\leq 26\%$ were assigned to the clinical cluster. For subjects performing above 26% correct on SRT_Tar, performance on NWRT_global was considered giving rise to two roots: a) if a subject performs $> 26\%$ on SRT_Tar and $> 60.61\%$ on NWRT_global, then assign to non-clinical cluster (TD), b) if a subject performs $> 26\%$ on SRT_Tar but $\leq 60.61\%$ on NWRT_global, then assign to clinical cluster (SLI).

¹⁶⁶ The classification accuracy given in squares within the ctree refers only to that of PAM clustering and does not represent diagnostic accuracy of cluster membership as measured by comparing cases identified as SLI or TD by the PAM to the clinical status assigned based on our classification procedure outlined in 3.1.

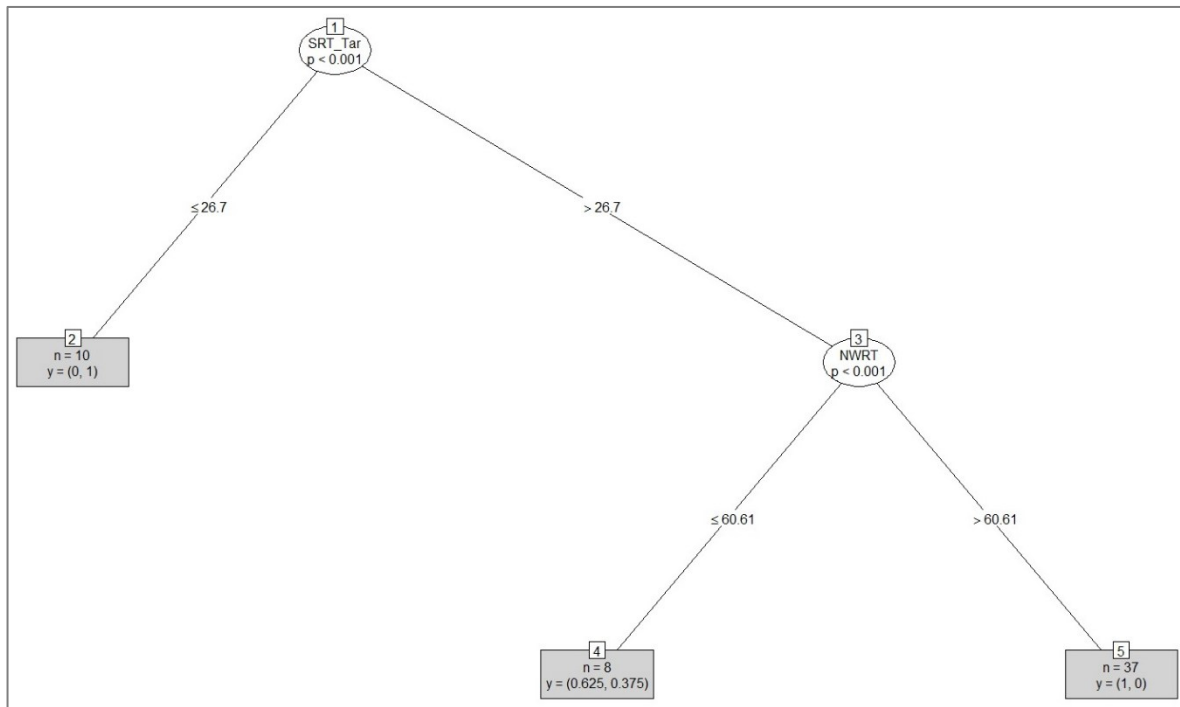


Figure 19: Conditional Inference Tree Analysis of the MoSLI + BiTD clustering result using SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD as predictors.

Turning to the results of bilingual children, we performed the PAM analysis on all BiSLI and BiTD groups collapsed together based on the performance scores in the SRT and NWRT. The *Hopkins statistic* indicated regularly spaced data that are neither clustered nor random (H-value of around 0.2), and the Gap statistic suggested the two-cluster solution. Results of the PAM clustering based on the performance of BiTDs and BiSLIs on SRT_Id and SRT_Tar were similar to those we obtained for BiTDs and MoSLIs (see **Figure 18**). 11/12 BiSLIs were assigned to the clinical cluster, yielding a sensitivity of 91.7 %, whereas 37/44 BiTDs were assigned to the non-clinical cluster giving a specificity of 84.1%, $LR^+ = 5.76$, $LR^- = 0.10$. The thresholds separating the bilingual clinical cluster from the non-clinical one were 33.3% for SRT_Id and 53.3% for SRT_Tar, whereby SRT_Tar was the main predictor for the clustering result (with the same cutoff of 53.3%). Regression analysis and ctree analysis showed that Positive_Early_Development (Firth: $\beta = 1.0636$, $Z = 2.614$, $p = .001$), followed by SES ($\beta = 0.7843$, $Z = 2.033$, $p = .01$), were significant predictors for cluster membership. Variables related to bilingualism, i.e., AoO, LoE, and LDI, did not explain cluster membership. An illustration of the hierarchical structure of variable importance with classification thresholds is depicted in **Figure 20**.

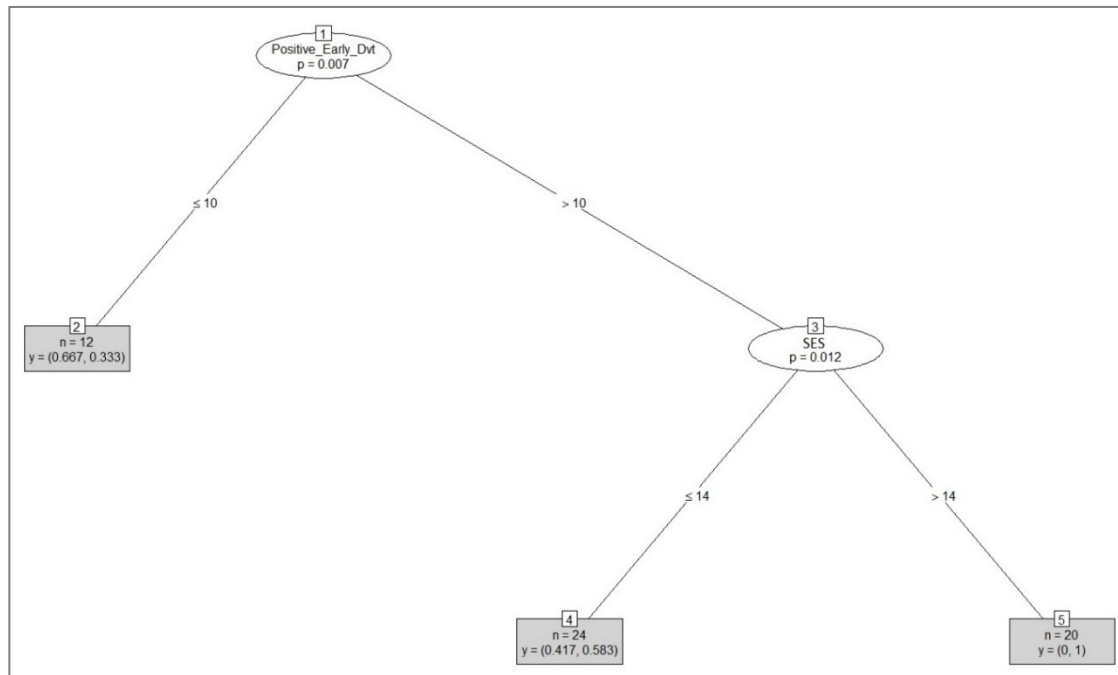


Figure 20: Conditional Inference Tree Analysis of the BiSLI+BiTD clustering result based on SRT_Id and SRT_Tar: hierarchical variable importance of background variables predicting cluster membership.

We ran the same clustering procedure on the bilinguals' performance in NWRT using the variables NWRT_global, NWRT_LI, and NWRT_LD. All BiSLI children were assigned to the clinical cluster yielding a 100% sensitivity; however, 9 BiTD children were assigned to the clinical cluster, i.e., only 35/44 BiTDs were assigned to the non-clinical cluster (specificity = 80%), LR+ = 5, LR- = 0. Ctree analysis showed that NWRT_global was the primary variable predicting cluster membership with a cutoff of 66.7%. Next, we ran Firth's biased regression analysis on the clustering results for NWRT entering age, Positive_Early_Development, SES, and LDI as fixed factors. Results showed that neither language dominance nor SES explained cluster membership based on NWRT_global. As expected, Positive_Early_Development was the main variable explaining the clustering result (Firth: $\beta = 0.38996$, $Z = 2.626$, $p = .001$). The other significant predictor for NWRT_global was chronological age (Firth: $\beta = 0.05931$, $Z = 2.150$, $p = .01$). Since NWRT_global is a composite score obtained by adding up scores of both the language-independent (NWRT_LI) and language-dependent parts (NWRT_LD), we wanted to verify whether both of them were affected by the age factor. To achieve this, we ran the PAM analysis on each of them separately. The results show that if clustering is solely based on performance on NWRT_LI upon a threshold of 73.3% (as established by ctree analysis), 10 BiTD children would be over-identified as having

SLI, yielding a specificity of only 77%, $LR+ = 4.385$, $LR- = 0.00$. Both Positive_Early_Development (Firth: $\beta = 0.38996$, $Z = 2.626$, $p = .01$) and age (Firth: $\beta = 0.05591$, $Z = 2.266$, $p = .01$) were significant predictors for the clustering results (variables entered in the regression model: Positive_Early_Development, SES, LDI, and age). Ctree analysis showed that the age threshold separating the two clusters based on NWRT_LI was 87 months (7;3 yrs.). On the other hand, if the bilingual children in our data set are clustered based on performance in NWRT_LD alone, the diagnostic accuracy drastically improves: upon a 50% cutoff score, only 2/44 BiTD children are assigned to the clinical cluster, while all BiSLI children are classified as SLI, which yields 95% specificity and 100% sensitivity ($LR+ = 20$, $LR- = 0.00$). Positive_Early_Development was singled out as a predictor explaining cluster membership based on NWRT_LD (Firth: $\beta = 0.30611$, $Z = 2.946$, $p = .001$), i.e., the variables age, LDI, and SES did not explain cluster membership. In the following step, we included all SRT and NWRT measures (SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD) in the clustering procedure. As can be seen in **Figure 21**, combining SRT with NWRT enhances diagnostic accuracy: all of the BiSLI children (12/12) were assigned to the clinical cluster (100% sensitivity), while 39/45 BiTD children were assigned to the non-clinical cluster (87% specificity, $LR+ = 7.692$, $LR- = 0.00$).

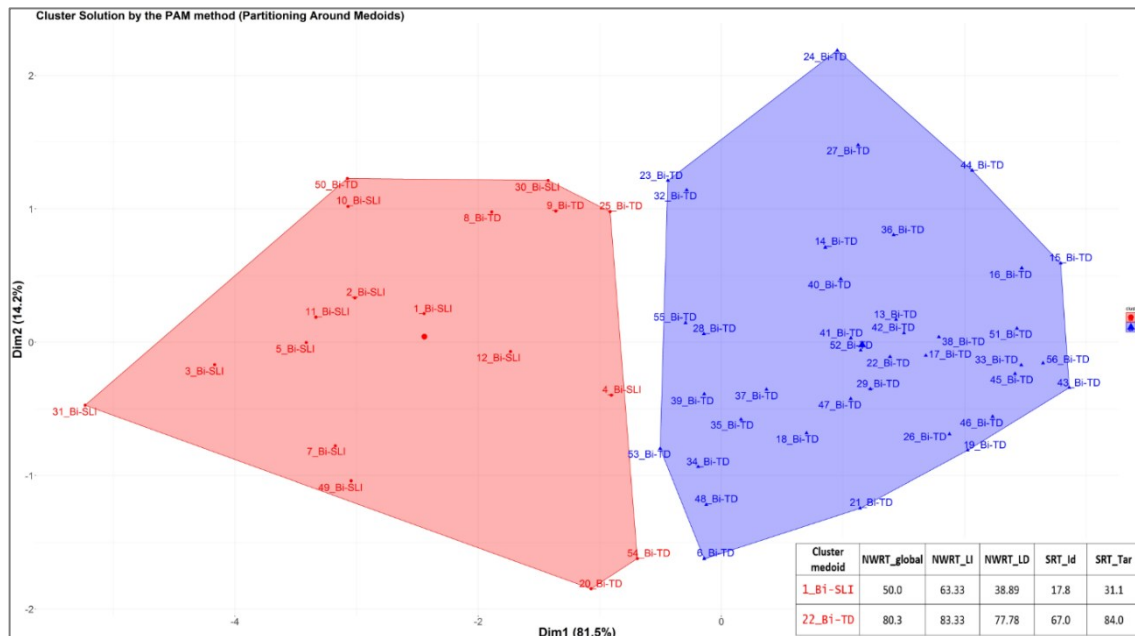


Figure 21: PAM-clustering of BiSLI+BiTD based on performance scores on SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD.

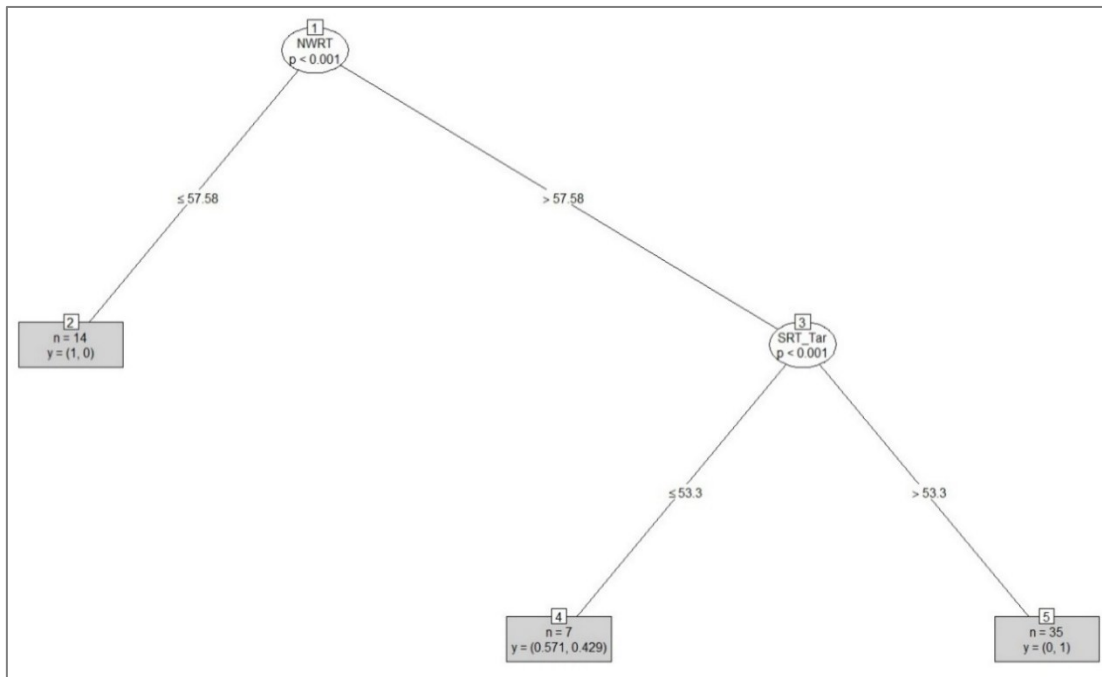


Figure 22: Conditional Inference Tree Analysis of the BiSLI+BiTD clustering result using SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD as predictors

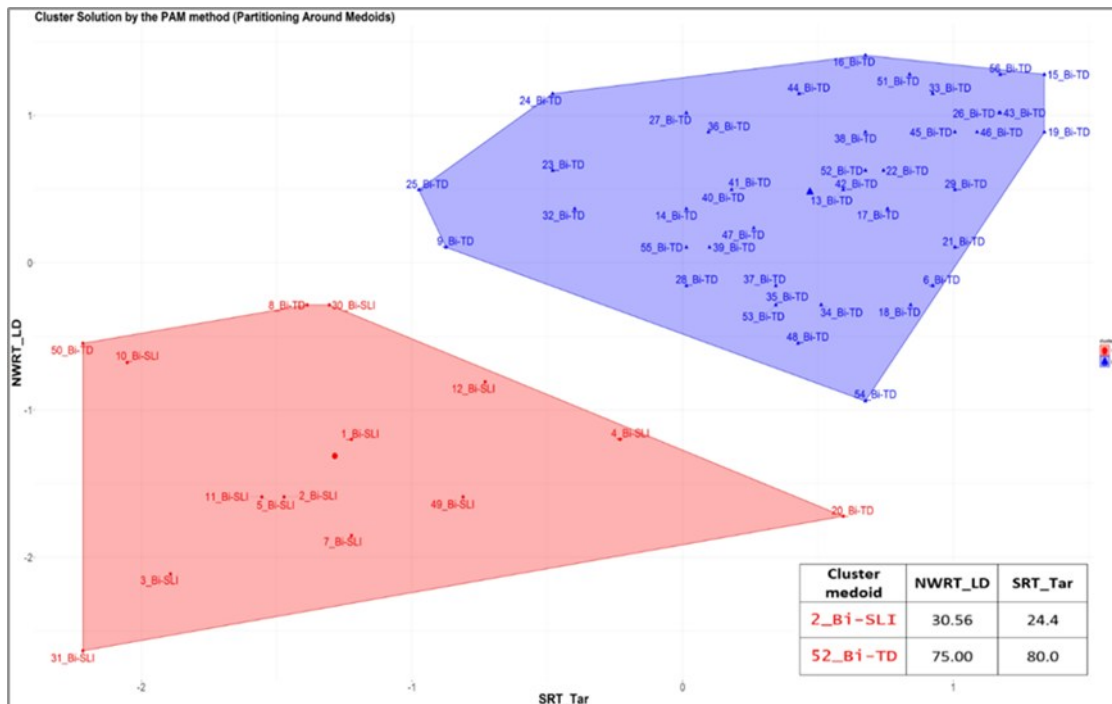


Figure 23: PAM-clustering of BiSLI+BiTD using performance scores on SRT_Tar and NWRT_LD

5.5 Discussion and conclusion

This study aimed to evaluate the robustness of two LITMUS tools, the German LITMUS sentence and quasi-universal nonword repetition tasks, against the influence of language dominance on their diagnostic accuracy for SLI in bilingual children aged 5;6-9;0 years. Since both tasks were designed to minimize bias against bilingual populations while being indicative of the presence of language impairment, we wanted to specifically verify whether the tasks were only sensitive to risk factors for SLI or whether background variables related to bilingualism, particularly the degree of language dominance (as estimated by the relative amount of use and exposure to L1/L2) could influence the performance of BiTDs to such an extent that would compromise the diagnostic accuracy of the tasks. The second aim of the study was to investigate whether combining LITMUS-SRT (especially when scored by correct target structure) with NWRT yielded better diagnostic accuracy than single measures and helped avoid cases of misdiagnosis. Following our own research (e.g., Abed Ibrahim & Hamann, 2017; Grimm & Hübner, forthcoming; Hamann & Abed Ibrahim, 2017; Hamann et al., 2017), we particularly wanted to check whether a combination of German SRT_Tar and the language-dependent part of the NWRT yielded higher diagnostic accuracy for identifying SLI in bilingual children than other combinations of measures. The former was found to be a fairer method than identical repetition for scoring SRT as it compensates for typical L2-errors such as lexical substitutions, while the latter was shown to maximize the performance gap between SLI and TD, not only in monolinguals but also in bilinguals, given its higher level of structural complexity.

In order to examine this, we first compared the overall performance of L1-dominant BiTDs to that of MoTDs, balanced and L2-dominant BiTDs, as well as to MoSLIs and BiSLIs. Results showed that although all three BiTD groups (regardless of their dominance profiles) significantly outperformed MoSLIs and BiSLIs on all SRT and NWRT measures, L1-dominant-BiTDs were significantly outperformed by MoTDs and L2-dominant-BiTDs on both SRT_Id and SRT_Tar with large effect sizes (see **Figures 15 and 16**). This echoes the findings of Meir (2018), who reported similar results for the performance of Russian-Hebrew bilinguals on LITMUS-SRTs in their weaker heritage or societal language. The results further showed that the performance gap between monolingual and bilingual SLI and TD groups was larger for NWRT_LD as opposed to the structurally less complex language-independent part of the LITMUS QU-NWRT and

the composite score of the two parts ‘NWRT_global’ (see **Figure 15**). This is in line with previous work showing that the complexity factors involved in the NWRT_LD part (i.e., presence of trilateral onset clusters and /sC/ clusters violating the Sonority Sequencing Principle) are particularly challenging for language-impaired children regardless of their lingual status (dos Santos & Ferré, 2018; Ferré et al., 2015; Grimm, 2022; Grimm & Hübner, forthcoming).

Since language dominance was used as a categorical variable to classify BiTDs in the between-group comparisons, one had to entertain the possibility that the assumed dominance effect for L1-dominant children might have been caused by confounding variables such as the age of onset of exposure to L2 (AoO), length of exposure to L2 (LoE), and socioeconomic status (SES). As for SRT_Id and SRT_Tar, moderate correlations were found between performance, language dominance, and SES, in addition to weak correlations for SRT_Id with LoE and AoO. Regression analysis showed that language dominance was the key predictor explaining variance in the performance of the BiTDs on SRT_Id and SRT_Tar, followed by SES. That AoO and LoE did not predict the performance of BiTDs on the SRT was an expected outcome since the vast majority of the participants in the bilingual sample were either simultaneous or early successive bilinguals and were exposed to German for at least 24 months at the time of assessment (see also Armon-Lotem et al., 2011, for similar results on L2-Hebrew-SRT).

The finding that language dominance influenced the performance of BiTDs on both measures of the LITMUS-SRT questioned its applicability for identifying SLI in L1-dominant children when administered in their weaker language (German). To answer this, a well-known unsupervised machine learning technique, the Partitioning Around Medoids (PAM), was used for establishing an automatic classification of the monolingual and bilingual children in the data set as TD vs. SLI directly from their performance scores on SRT and NWRT without using information about their clinical status. Subsequently, the participants’ clinical group membership revealed by PAM-clustering was compared to their clinical status based on standardized assessment in L1/L2, and sensitivity and specificity (diagnostic accuracy) levels of the tasks in isolation and combined were calculated. We also explored which combinations of the measures obtained from SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD yielded the highest diagnostic accuracy. Finally, we conducted regression analyses to investigate whether background variables other than risk factors for SLI, particularly language dominance (LDI), explained PAM-cluster membership as TD or SLI based on performance scores on SRT

and/or NWRT. Since the index of `Positive_Early_Development` was shown to be a strong predictor for SLI in bilinguals (Boerma & Blom, 2017; Tuller et al., 2018), our premise was that if PAM-cluster membership can only be predicted by this index and not by language dominance or other background variables known to influence performance on repetition tasks (age, AoO, LoE, SES), then clustering of cases cuts across the SLI/TD dimension confirming that the LITMUS-SRT and NWRT are only sensitive to the presence of SLI and are not biased against bilingual children who are non-dominant in the societal language.

In Hamann and Abed Ibrahim (2017), unsupervised (clustering) machine learning algorithms were only applied to the bilingual data, while the Receiver Operating Characteristic curve (ROC) analysis was used to calculate sensitivity and specificity levels for the monolingual data. Given that ROC analysis uses the predefined ‘clinical status’ (as assigned by standardized test procedures) as a dependent variable for predicting the sensitivity and specificity of a test, we wanted to verify this finding for the monolinguals using a method independent of ‘clinical status’. PAM-clustering solely based on scores in `SRT_Id`, `SRT_Tar`, `NWRT_global`, `NWRT_LI`, and `NWRT_LD` yielded even higher diagnostic accuracy than that in Hamann and Abed Ibrahim (2017). The fact that all of the subjects identified as MoSLI by standardized assessment belonged to the lower performing cluster, while all of the MoTDs belonged to the higher performing cluster (100% sensitivity and 100% specificity), provides additional evidence that these linguistically motivated tasks are very sensitive to the presence of language impairment in monolinguals and tap the core morphosyntactic and phonological deficits in SLI. Compared to results based on ROC-analysis in Hamann and Abed Ibrahim (2017), the source of the improved diagnostic accuracy is most likely the simultaneous inclusion of both tasks into the clustering procedure and the lower cutoff points obtained by applying ctrees to the PAM clustering. This is reminiscent of Armon-Lotem and Meir’s (2016) study, which reported an increase in diagnostic accuracy when LITMUS-SRT is supplemented by NWRT for Hebrew and Russian monolinguals. A further significant result was that chronological age could not predict cluster membership for the age range in our monolingual data set.

After establishing that LITMUS-SRT and LITMUS-QU-NWRT were sensitive to SLI in monolinguals, we addressed the frequently reported overlap between MoSLI and BiTD children (e.g., Armon-Lotem, 2010; Håkansson & Nettelbladt, 1996; Hamann, 2012; Paradis, 2010a, 2010b). PAM-clustering conducted on SRT scores entering both

measures, SRT_Id and SRT_Tar, yielded good overall diagnostic accuracy (100% sensitivity and 84.1%¹⁶⁷ specificity), with SRT_Tar being the leading variable for predicting cluster membership since it led to a better separation between the BiTD and MoSLI clusters. Several studies found this scoring method better suited for assessing morphosyntactic abilities in bilingual children since it only focuses on the mastery of syntactic structure and does not penalize bilingual children for frequent L2-errors such as lexical substitutions (Abed Ibrahim et al., 2018; Armon-Lotem & Meir, 2016; Hamann & Abed Ibrahim, 2017; Hamann et al., 2017; Meir, 2018).

Next, we checked whether the overlap problem between MoSLI and BiTD could be overcome using SRT combined with QU-NWRT. Indeed, including NWRT scores in the clustering procedure resulted in a much better diagnostic accuracy with almost no overlap between MoSLI and BiTD (100% sensitivity and 95% specificity). As also reported in Armon-Lotem & Meir (2016), de Almeida et al. (2017), Boerma & Blom (2017), and Hamann & Abed Ibrahim (2017), the latter finding corroborates that a combination of LITMUS instruments assessing different areas of language ability helps avoid cases of misdiagnosis. Among the five measures, SRT_Id, SRT_Tar, NWRT_global, NWRT_LI, and NWRT_LD, both SRT_Tar and NWRT_global were the main predictors for clustering results, with SRT_Tar being the more important contributor (see **Figure 19**). We further demonstrated that chronological age did not predict cluster membership here either.

Regarding the diagnosis of bilinguals, PAM clustering based on scores in SRT_Id and SRT_Tar resulted in good overall accuracy rates (91.7% sensitivity and 84.1% specificity). Interestingly, the same 7 BiTDs previously assigned to the clinical cluster upon comparison with MoSLIs were also classified as SLI by the PAM, showing that changing the reference group did not influence the individual classification of the BiTDs. Again, SRT_Tar, which compensates for L2-errors, was the primary contributor to the clustering results with a cutoff of 53.3%, which is very close to the threshold obtained by *k*-means clustering in Hamann and Abed Ibrahim (2017). Of the five background variables considered for regression analysis, just two variables unrelated to bilingualism

¹⁶⁷ It is worthwhile mentioning that three of the 7 BiTD children assigned to the clinical cluster started acquiring L2 German after the age of five and spent their first years in an exclusive L1-environment, demonstrate high rates of determiner errors, especially case and gender errors which could affect the realization of target structure of a sizable proportion of the test items of the LITMUS-SRT (see also Abed Ibrahim et al., 2018 and Chilla et al., 2021).

emerged as significant predictors for clustering membership: Positive_Early_Development, followed by SES. The influence of language dominance, which was a significant predictor for explaining the variance in the performance of the BiTDs in SRT_Id and SRT_Tar, was outweighed by the presence of risk factors for SLI and was rendered insignificant once the BiSLIs became part of the equation. This is consistent with the findings of Tuller et al. (2018), who found for the German sample that Positive_Early_Development was the leading predictor for performance in SRT (followed by SES) over variables related to bilingualism.

The clustering solution based on NWRT_global, NWRT_LI, and NWRT_LD scores yielded only fair diagnostic accuracy rates due to reduced specificity (specificity =80%). NWRT_global emerged as the main predictor for clustering results. Regression analysis revealed that Positive_Early_Development (the most important predictor) and chronological age were significant predictors for clustering results based on performance scores in NWRT_global. Given that NWRT_global is a composite score computed by adding up performance scores in NWRT_LI and NWRT_LD, and since Grimm (2022), as well as Grimm and Hübner (forthcoming), reported an overlap between MoSLI and BiTD on NWRT_LI and better discriminatory power for NWRT_LD in children aged 8;0 to 10;0 yrs., we ran cluster analyses on both subparts of the NWRT separately to check for age effects. The analysis revealed that, in addition to Positive_Early_Development, cluster-membership based on NWRT_LI was predicted by chronological age with a threshold of 7;3 years, whereas cluster-membership based on NWRT_LD was not predicted by age and was only sensitive to risk factors for SLI. On the other hand, neither bilingualism-related factors nor SES predicted cluster membership derived from performance scores on NWRT_global or subtests thereof. The latter result echoes what has been found for this type of NWRT in de Almeida et al. (2017) and Tuller et al. (2018).

We have also shown that including all five SRT and NWRT measures in the clustering procedure enhances diagnostic accuracy for SLI in bilingual children, where NWRT_global and SRT_Tar were the main contributors explaining the results of the cluster solution. Interestingly, once SRT is combined with NWRT, only Positive_Early_Development emerges as a significant predictor for clustering results, and SES does not play a role anymore, which is in line with the findings of Chiat and Polišenská (2016).

Given that clustering by scores on NWRT_LI appeared to be influenced by age, while NWRT_LD was only sensitive to risk factors of SLI

(Positive_Early_Development), and since SRT_Tar was the chief contributor towards clustering results when both SRT_Id and SRT_Tar are included in any clustering procedure on bilingual performance, we expected a combination of SRT_Tar and NWRT_LD to yield better diagnostic accuracy rates than other combinations of measures. Indeed, clustering based on performance scores in SRT_Tar and NWRT_LD yielded the highest classification accuracy, where only Positive_Early_Development predicted clustering results. The crucial contribution of the structurally more complex NWRT_LD towards diagnostic accuracy is consistent with the robust effects of phonological complexity found in the respective studies (e.g., Gallon et al., 2007; Ferré et al., 2012), with clinical implications that phonological complexity can be used as a reliable indicator for SLI in both monolingual and bilingual children (see de Almeida et al., 2019; Grimm, 2022; Grimm & Hübner, forthcoming). Our results concerning the NWRT_LD part might seem at odds with findings of other studies showing better diagnostic accuracy for *Crosslinguistic*-NWRTs over *Language-Specific*-NWRTs in bilingual populations, e.g., Armon-Lotem and Meir (2016), Boerma and Blom (2017), Boerma et al. (2015), Öberg and Bohnacker (2022). This can be ascribed to differences in the construction of the tasks, which, as described in 2.3.1.1 and 2.3.1.2, tap different aspects vulnerable in SLI (i.e., phonological working memory vs. phonological complexity) and differ considerably from each other, especially in their language-dependent parts. Another possible reason for the poor diagnostic accuracy reported for the *Language-Specific*-NWRTs in the latter four studies might be the relatively young age of their participants (5;0-6;0 in the first three studies and 4;0-7;11 in Öberg and Bohnacker's study) compared to the age range in the sample of this study (5;6-9;0), which covers the last year of preschool and the first 2-3 primary school years. A study by Rispen and Baker (2012) demonstrated that both lexical knowledge and discrimination ability significantly influenced performance on NWRT in 5-year-old MoTDs, while this kind of relationship could not be attested for 8-year-olds, see also Öberg and Bohnacker (2022) for similar findings on the interrelation between lexicon and NWRT in bilinguals.

In line with our previous research, the results presented here and the fact that they emerge from unsupervised PAM-clustering indicate clearly that the German LITMUS-SRT and QU-NWRT are promising tools for the identification of language impairment in bilingual populations with diverse language dominance profiles. The study replicated the finding that SRT_Tar is better suited than SRT_Id for assessing the language abilities of bilingual children with German as L2 on a slightly larger group of children with a

statistical method better suited to the study's data set. Even though dominance influences the performance of BiTDs, especially in the SRT, we demonstrated that the diagnostic accuracy of these tools is not compromised by language dominance. While risk factors for SLI were significant predictors for clinical status in all models, language dominance did not contribute at all to explaining the results of any of the clustering procedures. Moreover, our results confirmed that using a combination of tasks, each emphasizing a different aspect of language ability, enhances diagnostic accuracy and helps avoid cases of misdiagnosis. As a last promising result, we showed that using SRT_Tar in conjunction with NWRT_LD renders the best diagnostic accuracy so far obtained in studies on similarly constructed tasks, where the combination of measures is only sensitive to risk factors for SLI but not to language dominance nor to SES, which is not achieved by many tasks. We, therefore, feel confident in pursuing these investigations in order to be able to provide reliable and easy-to-administer L2 tools for clinical use in bilingual contexts.

6 Study 3: German LITMUS sentence-repetition task as a screening tool for bilingual children in Germany: The role of syntactic complexity and working memory¹⁶⁸

Abstract

Given the cross-linguistic vulnerability of complex syntax in children with SLI and its relative robustness in bilingual typical language development, computationally complex constructions were incorporated into the German LITMUS-sentence-repetition-task (SRT, Hamann et al., 2013). Such constructions are particularly challenging for children with SLI given their impaired language faculty associated with weaker/less stable linguistic representations in long-term memory (LTM) and limited working memory (WM) capacity (Jakubowicz, 2005, 2011).

This study investigates how syntactic complexity interacts with repetition accuracy (scored by correct target structure) in 106 children with and without SLI (40 monolinguals and 66 bilinguals, ages 5;6-9;2). To achieve this, the sentences of the German LITMUS-SRT were grouped into different complexity levels depending on the number of syntactic operations (syntactic movement, clausal embedding) involved in the derivation of the structure, and intervention (NP-restriction) as an additional complexity factor (Jakubowicz, 2011; Rizzi, 2004). Regression analyses were run to determine whether monolingual and bilingual children with and without SLI demonstrated different patterns of association between performance on complex constructions and linguistic and WM measures. Verbal short-term memory was indexed by forward digit span (FDS) and verbal working memory by backward digit span (BDS).

In line with previous studies (Delage & Frauenfelder, 2019, 2020; Friedmann & Reznick, 2021), repetition accuracy decreased in all groups as a function of cumulative syntactic complexity, with larger effect sizes for the presence of two or more complexity factors in the SLI groups. An interaction was observed between age and complex

168 Parts of this chapter were published in the Proceedings of the 48th Annual Boston University Conference on Language Development after the thesis defense. Full reference: Abed Ibrahim, Lina & Hamann, Cornelia (2024). German Litmus Sentence-Repetition Task as a screening tool for the identification of SLI in bilingual children in Germany: The role of syntactic complexity and working memory. In H. A. A. AlThagafi & J. Ray (Eds.), *Proceedings of BUCLD 48* (pp. 1-16). Somerville: Cascadia Press.

structures in the BiTD group: younger BiTD children (5;6-6;11) performed worse than MoTDs on structures involving A'-movement, in which theta-role assignment requires exploiting case marking. Despite not showing mastery, BiTDs still outperformed children with SLI, who showed near-floor performance on structures involving two or more complexity factors.

No associations with WM emerged for the SLI groups, whose performance was predicted by L2-morphosyntax. This seems at odds with previous studies (Delage & Frauenfelder, 2020; Zebib et al., 2019). However, within-group distributions of FDS and BDS in the SLI data revealed that most children had a span of 3 on FDS and almost floor performance on BDS, reducing the inter-individual variability and thus an association with performance. As to the TD children, performance of MoTDs on Level 1 structures was predicted only by vocabulary measures, whereas BiTDs' performance was predicted by FDS and L2-vocabulary. Interestingly, performance on the structurally more complex Level 2 was only predicted by FDS in both TD groups confirming that processing of computationally more complex structures necessitates more WM capacities (Arosio et al., 2011; Jakubowicz, 2011). BiTDs relying more on WM for processing complex constructions than MoTDs might be the result of poor L2-vocabulary, a notoriously vulnerable domain for bilinguals. While MoTDs can rely on their more advanced language skills (i.e., language knowledge in LTM), BiTDs compensate for weaker (L2) language abilities by resorting to their intact WM capacities. The findings of the study are of great relevance not only for theories of the causes of SLI, but also for shortening and ameliorating the task to ensure maximal diagnostic accuracy in bilingual populations.

6.1 Introduction

As demonstrated in Chapter 2 (see 2.1.2.2 & 2.2.1.1), complex syntax constitutes a relative area of strength in bilingual typical language development, being more robust to sources of individual variation than other language domains, such as vocabulary and inflectional morphology. At the same time, it is a vulnerable area for children with SLI, who have consistently been found to experience pronounced difficulties in the production and comprehension of complex (morpho)syntactic constructions (e.g., Friedmann & Novogrodsky, 2011; Hamann et al., 2007, 2017; Hamann & Tuller, 2014; Jakubowicz & Tuller, 2008; van der Lely, 1998). In particular, it has been shown that clausal embedding is preserved in BiTD children, whereas it is difficult for monolingual and bilingual children with SLI (Delage et al., 2007; Scheidnes & Tuller, 2019). The same is true for *wh*-movement, which is relevant for (German) constituent questions, relative clauses, and topicalization (Biran & Ruigendijk, 2015; Jakubowicz, 2011; Prévost et al., 2014; Ruigendijk & Friedmann, 2017). In question formation and relative clauses, an additional factor identified as difficult is intervention, specifically, the occurrence of a lexical DP between the *wh*- or relative element and the subject of the relative clause as it occurs in object-which questions and some object relative clauses (Friedmann et al., 2009; Hamann & Tuller, 2015; Rizzi, 2018; see 2.1.2.2 and 2.2.1.1 for a detailed description). Another factor often assumed to increase difficulty for children with SLI is the so-called non-canonical word order, which makes such sentences harder to parse because thematic role assignment cannot proceed from the base-generated order and instead requires reconstruction of non-local dependencies, exploiting morphological cues in addition to word-order cues (Adani et al., 2010, 2014). Examples of constructions with non-canonical word orders include object questions, object relatives, and passives (see van der Lely 1996, Ruigendijk & Friedmann, 2017).

It is assumed that computationally complex constructions interact with working memory capacities, especially when the processing load increases due to the co-presence of multiple complexity factors (Adani, 2011; Arosio et al., 2011; Grillo, 2008; Ruigendijk & Friedmann, 2017; for a detailed description of the different WM components, see 2.1.3). This holds for clausal embedding as well as for filler-gap constructions created by syntactic movement, especially for those with interveners (see Gibson, 1998; Gordon et al., 2001; Kimball, 1973; but see Pickering & van Gompel, 2006, for a different

approach). The relatively good performance of BiTD children is therefore attributed to the fact that they have an intact language faculty and working memory. Thus, having been acquired in their first language, syntactic operations such as recursion (de Villiers et al., 2020), clausal embedding (Delage et al., 2008; Hamann et al., 2007; Hamann & Tuller, 2014), and (*wh*-)movement (Biran & Ruigendijk, 2015; Jakubowicz, 2011; Prévost et al., 2014) do not have to be reacquired and should be available for their L2-grammar given sufficient exposure to the respective language (Hamann et al., 2017; Roeper, 2011; Schwartz, 2004; see also Scheidnes, 2020). Children with SLI, on the other hand, are argued to have deficits in linguistic representations or derivational operations in addition to limited WM capacity. Consequently, they are expected to struggle disproportionately with such complex syntactic computations (Friedmann et al., 2009; Hamann et al., 2017). For this reason, computationally complex constructions were incorporated into the German LITMUS-sentence-repetition-task (SRT, Hamann et al., 2013, 2017), with special attention to clausal embedding, *wh*-movement, intervention, and non-canonical word order.

The interaction between WM capacities and the comprehension and production of complex constructions has been established for typically developing and language-impaired children (e.g., Adani, 2011; Arosio et al., 2011; Bentea & Durrleman, 2014; Bentea et al., 2016; Delage & Frauenfelder, 2019, 2020; Dispaldro et al., 2011; Grüter & Crago, 2012; Willis & Gathercole, 2001). For example, Arosio et al. (2011) found that VSTM, indexed by forward digit span (FDS), modulated performance on comprehension of object relative clauses in Italian-speaking MoTD children (mean age = 9;3). Similarly, Bentea and Durrleman (2014) found that comprehension of *wh*-questions and relative clauses in 5-year-old MoTD French-speaking children was linked to their WM abilities (measured by forward and backward digit span). Interestingly, WM effects were much weaker for the older groups in their study (age range: 7;0-9;0) and were confined to very complex constructions with intervention effects, e.g., the object relative “*Montre-moi le chat que le garçon lave*” (Show me **the cat** that **the boy** is washing) (Bentea & Durrleman, 2014, p. 86). Further studies found a link between verbal complex memory tasks assessing verbal working memory (e.g., backward digit span) and comprehension of complex constructions in typically developing children (e.g., Engel de Abreu et al., 2011; Montgomery et al., 2008). In a longitudinal study, Verhagen and Leseman (2016) found that VWM abilities at ages 4-5 significantly predicted comprehension of

morphosyntax two years later in typically developing Dutch-speaking monolingual and bilingual children. Delage and Frauenfelder (2019) looked into the relationship between working memory and processing of complex syntax (sentence comprehension, sentence repetition, and spontaneous production) in 48 typically developing French-speaking children aged 5 to 12. Their results showed that, in line with Jakubowicz (2011), both WM and complex syntax significantly improved with age. Importantly, they found a strong link between WM (measured by complex span tasks) and performance on all three measures. Analysis of the sentence repetition task revealed that the VWM effects were modulated by the degree of computational complexity: complex span tasks better predicted performance on more complex structures relative to less complex ones, i.e., deeply embedded clauses (at least two levels of embedding) rather than less deeply embedded clauses and object relatives rather than subject relatives. In the same vein, Schouwenaars et al. (2018) investigated the comprehension of which-questions (subject which-questions, object which-questions, and passive-which questions) by L1 German-speaking children (ages 7-9) and adults using a picture selection task with eye tracking. Offline accuracy scores showed that children performed significantly more poorly on object which-questions than on subject and passive which-questions, and that performance was influenced by digit span (raw span of backward digit span) and age. Children with a span of 3 (low) demonstrated lower accuracy scores than children with medium (span = 4) or high BDS spans (span = 5-6), and 7-year-olds were outperformed by their 8- and 9-year-old counterparts. Moreover, their online gaze data revealed that children with higher WM capacities and adults were more inclined to use morphosyntactic cues (case and number agreement between subject and verb) for interpreting object which-questions than children with lower WM capacities.

Studies featuring children with SLI also showed associations between WM components and complex syntax, though there were inconsistent findings as to which WM component better explained variance, namely verbal short-term memory (VSTM) as measured by simple span tasks or VWM as indexed by performance on complex span tasks. For example, Delage and Frauenfelder (2020) examined the link between WM and the processing of complex sentences (in spontaneous production and comprehension and sentence repetition) in French-speaking monolingual children with ($n = 28$; age range: 5;0-14;6, mean: 8;10) and without SLI ($n = 48$, age range: 5;2- 12;9, mean: 9;0). A general finding of the study was that working memory capacities and syntactic complexity

significantly increased with age in the TD children, whereas the progression for the SLI group appeared to be slower (note that this is a cross-sectional study). Consistent with previous findings, children with SLI showed substantial deficits in both WM and complex syntax. Interestingly, complex constructions were predicted by different WM components in the TD and SLI groups in spontaneous production (as measured by MLU, rate of embedded clauses, and rate of multiple embeddings). Simple structures were not correlated with any WM measures, whereas complex constructions were predicted by complex spans (in particular, the counting span) in the TD group and only by simple spans in the SLI group. Concerning comprehension and repetition of complex structures, a significant proportion of the variance in comprehension was accounted for by a combination of age, simple, and complex spans in both TD and SLI groups. Repetition of complex constructions was predicted by age, simple and complex span tasks in the TD group, while only simple and complex span tasks predicted performance in the SLI group, with only a marginally significant effect for chronological age. This implies that SLI children's WM capacities deviate in their maturational schedule from those of typically developing children.

Zebib et al. (2020b)¹⁶⁹ examined the effects of WM and overall language abilities on the performance of bilingual typically developing and language-impaired children on the French LITMUS-SR (Fleckstein et al., 2018), which is constructed by the same principles as the German LITMUS-SRT. The study was conducted on a large sample of bilingual children aged between 5;6 and 8;1, with the same 3 L1s as in the current study. A crucial finding of their study was that language measures and WM components differently predicted the performance of the BiTD and BiSLI children. Performance of the BiTD children was mainly explained by language measures, followed by a smaller contribution of VSTM (indexed by forward digit span), while performance in the BiSLI group was predicted by VWM (indexed by backward digit span), with only phonology explaining an additional portion of the variance. The authors argue that BiSLI children highly rely on their VWM abilities due to their impoverished linguistic representations, whereas BiTD children profit from their better linguistic representations, which are

¹⁶⁹ Note that Zebib et al.'s (2020b) study initially intended to feature the bilingual German LITMUS-SRT data collected within the same project (BiLaD) in Germany. However, due to the small size of the German BiSLI sample at the time of the study, analyses were restricted to the French data.

assumed to play a facilitating role and alleviate the processing load in sentence recall. According to Riches (2020, p. 503), this might imply that “children with DLD/[SLI] use different mechanisms to perform the same task” as postulated in Pinker and Ullman’s (2002) dual-root account. That VWM but not VSTM predicted variance in the BiSLI group is noteworthy as it contradicts the findings of other studies, which either found no links of either WM measure to repetition in children with SLI (e.g., Riches, 2012) or found only VSTM effects (e.g., Delage & Frauenfelder, 2020), or reported both BiTD and BiSLI children to rely only on VWM (Talli & Stavrakaki, 2020).

6.1.1 The present study

Given their robust acquisition in bilingual typical language development and cross-linguistic vulnerability in children with SLI, complex structures involving embedding, syntactic movement, intervention (NP-restriction), or combinations thereof are expected to identify SLI in bilingual children with great accuracy. Hence, they were incorporated into the German LITMUS SRT, alongside structures that mark crucial milestones in the acquisition of German word order properties (see Section 3.2.3.1 for details). As shown in the studies discussed in the theoretical part of this thesis and in Study 2, the German LITMUS-SRT, especially when scored by correct target structure, reliably identifies SLI not only in monolinguals but also in bilinguals. Moreover, the diagnostic accuracy of the task is not compromised by factors related to bilingualism, particularly the degree of language dominance (though it contributes to the variance in performance of the BiTD children), especially when the task is combined with the structurally more complex language-dependent part of the German LITMUS QU-NWRT (Abed Ibrahim & Fekete, 2019; Grimm, 2022). Furthermore, the task appears to be relatively robust against cross-linguistic influence from the L1 (Abed Ibrahim et al., 2018; Chilla et al., 2021). However, what has not yet been explored for the German LITMUS-SRT is how syntactic complexity interacts with repetition accuracy in the different lingual (monolingual vs. bilingual) and clinical groups (TD vs. SLI) in a larger participant sample, and whether the variance of performance is primarily explained by language abilities or by WM abilities. Consequently, the study has two aims: (1) to determine whether the previously reported diagnostic accuracy replicates in a larger sample, and (2) to examine how this accuracy interacts with syntactic complexity. Specifically, the study investigates the effects of

(cumulative) syntactic complexity on the performance of monolingual and bilingual children with and without SLI. To this end, the sentences of the German LITMUS-SRT were grouped into four complexity levels; complexity for each sentence type was estimated by the number of morphosyntactic operations required for its derivation, with intervention (NP-restriction) included as an additional complexity factor. To complement prior work on the influence of working memory and linguistic abilities on performance across the different populations and constructions, the study also explores language measures (lexical and morphosyntactic abilities as estimated by standardized language tests; Klem et al., 2015) and WM measures (as estimated by simple and complex digit recall tasks: FDS and BDS) as potential predictors of performance. Against the background of the linguistic theory and empirical findings described above, the aims of the study are as follows:

- i. To estimate the diagnostic accuracy levels (sensitivity and specificity) of the German LITMUS-SRT based on a larger participant sample and to determine the cutoff points that yield the highest diagnostic accuracy for this sample.
- ii. To investigate how cumulative syntactic complexity, operationalized in terms of (*wh*-)movement, syntactic embedding, and (movement) intervention, interacts with repetition accuracy in monolingual and bilingual children with and without language impairment. Repetition accuracy is expected to decrease as a function of cumulative syntactic complexity. Moreover, complexity is hypothesized to affect performance differently in the TD and SLI groups, while the effects of bilingualism are expected to be negligible.
- iii. To examine whether monolingual and bilingual children with and without SLI demonstrate different patterns of association between SRT performance (overall and by complexity level) and other linguistic measures, as well as verbal short-term memory (VSTM) and verbal working memory (VWM).

6.2 Method

6.2.1 Materials

Given the risk of over- and underdiagnosis with SLI in multilingual contexts (Grimm & Schulz, 2014), all bilingual participants underwent an extensive screening procedure, which included standardized assessment in their L1 and L2 and evaluation of non-verbal intelligence following the experimental protocol outlined in **3.1.1**. The German LITMUS-SRT (Hamann et al., 2013) was administered to all children; for a detailed description of the test, see **3.2.3.1**. Additionally, to examine the influence of working memory (VSTM and VWM) on sentence repetition, the forward digit span (simple span: FDS as a measure of VSTM) and backward digit span (complex span: BDS as a measure of VWM) subtests from the WISC-IV (Petermann & Petermann, 2011) were administered to the children, see **3.2.2** for a detailed description. The Questionnaire for Parents of Bilingual Children (PaBiQ, Tuller, 2015) was used to gather background information about the children, including language exposure variables, parents' socioeconomic status, and risk factors for SLI (see **3.2.3.3**).

6.2.2 Participants

The study features 106¹⁷⁰ monolingual and bilingual children aged between 5;6 and 9;2 years: 17 monolingual typically-developing children (MoTD), 23 monolingual children with SLI (MoSLI), 47 bilingual typically-developing children (BiTD), and 19 bilingual children with SLI (BiSLI). The bilingual children had Arabic, Portuguese, or Turkish as L1s and were acquiring German either simultaneously with their L1 (36/66) or successively as their L2 upon kindergarten entry (30/66), and had at least 18 months of L2 exposure. All participating children, including those with SLI, had an IQ score of 85 and above according to Wechsler's IQ scale. A participant overview is provided in **Table 23**. Pairwise comparisons correcting for Type I error revealed no significant between-group differences between the bilingual groups on any of the variables presented in **Table 23**.

¹⁷⁰ The participant sample consists of the 77 children who participated in Study 2 and 29 children with and without language impairment who were recruited for the BiliSAT project.

Table 23: Participants including monolingual control groups after verification of clinical status: mean (SD) and range

	MoTD (<i>n</i> = 17)	MoSLI (<i>n</i> = 23)	BiTD (<i>n</i> = 47)	BiSLI (<i>n</i> = 19)
Simult./total			24/47	12/19
Age at testing	82.9 (13.5)	83.4 (10.2)	87.3 (12.9)	78.3 (10.7)
(in months)	66-105	68-103	66-110	64-103
Age of Onset	0	0	29.9 (24.8)	22.3 (17.6)
(AoO_L2 in months)			0-90	0-59
Length of Exposure	82.9 (13.5)	83.4 (10.2)	60.7 (25.8)	58.2 (19.5)
(LoE_L2 in months)	66-105	68-103	18-110	27-88
Socioecon. status (SES)	N/A	N/A	13.0 (4.8)	12.8 (2.9)
(yrs mater. educ.)			8-22	8-22
Gender	12 F, 5 M	10 F, 13 M	25 F, 22 M	4 F, 15 M

Notes. MoTD: monolingual typically developing; MoSLI: monolingual with language impairment; BiTD: bilingual typically developing; BiSLI: bilingual with language impairment.

6.2.3 Data analyses and statistical treatment

The children's responses on the SRT were audio-recorded, transcribed, verified, and scored offline following the procedure outlined in 3.2.3.1. Null reactions were counted as errors unless they arose due to technical issues (missing data, less than 1% of the overall data). Statistical analyses were conducted using IBM SPSS 27 (2020). For group comparisons, non-parametric tests were employed due to unequal sample sizes and violation of the normality assumption, as verified by the Shapiro-Wilk test. Diagnostic accuracy and cutoff points separating SLI from TD in monolinguals and bilinguals were determined for both scoring measures, "identical repetition" and "target structure", using receiver operating characteristic (ROC) analysis (Dunn, 2014), complemented with likelihood ratios (LRs), see 2.3.2 for details. The latter analyses enabled calculating sensitivity, i.e., true positive cases (the probability that a child with genuine language impairment will be identified as such by the SRT), and specificity, i.e., true negatives (the probability that a child with typical language development will be identified as such by the SRT). Correlational and hierarchical regression analyses were conducted to assess the influence of linguistic and WM variables on SRT performance. Since scoring sentences by "target structure" has repeatedly been shown to yield better diagnostic accuracy for

the German LITMUS-SRT than “identical repetition” (Abed Ibrahim & Fekete, 2019; Hamann & Abed Ibrahim, 2017), the rate of correctly realized target structure (henceforth SRT_Tar) was used as the basis for regression analyses and subsequent analyses of syntactic complexity effects.

To investigate the effects of syntactic complexity, the 45 sentences of the German LITMUS-SRT (15 conditions á 3 sentences) were grouped into four different complexity levels depending on the presence of complexity factors or combinations thereof. The complexity of a given sentence type was estimated by the number of relevant morphosyntactic operations involved in its derivation, in addition to the presence of intervention (NP-restriction as an additional complexity factor). First, the so-called ‘0-level’ sentences constitute the least complex constructions since they do not involve (*wh*-)movement, embedding, or intervention. This category comprises 21 sentences in 7 conditions, with three items per condition. The following sentence types are included in the 0-level group: intransitive sentences with the main verb ‘werden’, simple SVO sentences (3 present and 3 regular preterite forms), coordinated constructions¹⁷¹, and three types of structures that involve the German sentence bracket¹⁷². The latter are SA/OV sentences with participial constructions (modal + participial form of the lexical verb), constructions with auxiliary verbs + infinitives (non-finite complement clauses), and SVO sentences with particle verbs. Level-1 sentences include either (*wh*-)movement or embedding as a syntactic complexity factor, but not both. It consists of 9 sentences in 3 conditions: finite complement clauses (embedding), object bare *wh*-questions (*wh*-movement), and long passives (A-movement/smuggling¹⁷³). Level-2 comprises structures with two syntactic complexity factors and includes 12 sentences in 4 conditions: subject relatives (*wh*-movement and embedding), object relatives without intervention¹⁷⁴ (*wh*-movement and embedding), topicalized (ObjVS) constructions (*wh*-movement and

¹⁷¹ Coordinated constructions contained an equal or a higher number of syllables (13/14) than all other structures of the German LITMUS-SRT.

¹⁷² Such sentences contain a number of syntactic operations such as verb-movement but are, nonetheless, classified as 0 since they do not contain the crucial operations. The sentence bracket, a mile-stone in L2-acquisition, has been reported to be acquired fast and together with verb morphology.

¹⁷³ Following the smuggling analysis of Collins (2005), in passive constructions, the VP is first moved as a chunk and the object then moves out; consequently, intervention effects are avoided because the object no longer has to cross the subject (see Belletti & Rizzi, 2012).

¹⁷⁴ In order to keep memory load as constant as possible in the subject and object relatives without intervention, Hamann et al. (2013) used only two lexical DPs in these structures and the verb see consistently in the matrix clause.

intervention), and object which-NP questions (movement and intervention). Note that it is not necessarily assumed that the constructions within a given level are equally complex, e.g., subject relatives have been assumed to be computationally less complex than object relatives without intervention due to the canonical word order within the embedded clause in SRs, which preserves theta-role assignment. But see the results from Hamann et al. (2017) regarding this point. Finally, Level-3 consists of the most complex sentence type, namely object relative clauses with an intervening lexical DP (*wh*-movement, embedding, and intervention); see **Table 24** for an overview.

Table 24: Syntactic complexity levels based on movement, embedding, and intervention

		Movement	Embedding	Intervention
Level 0	SVO-present	(-)	(-)	(-)
	SVO-simple past	(-)	(-)	(-)
	Sentence bracket (Aux)	(-)	(-)	(-)
	Sentence bracket (particle)	(-)	(-)	(-)
	„Werden“ Control	(-)	(-)	(-)
	Non-finite complement clause	(-)	(-)	(-)
	Coordination	(-)	(-)	(-)
Level 1	Finite complement clauses	(-)	(+)	(-)
	Object who-question	(+)	(-)	(-)
	Long passive	(+) A-movement	(-)	(-)
Level 2	Topicalization (OVS)	(+)	(-)	(+)
	Object which-question	(+)	(-)	(+)
	Subject relative clause	(+)	(+)	(-)
	Object relative clause without intervener	(+)	(+)	(-)
Level 3	Object relative clause with intervener	(+)	(+)	(+)

6.3 Results

6.3.1 Overall performance on the German LITMUS SRT and diagnostic accuracy

First, Omnibus Kruskal–Wallis tests were run using the overall SRT scores scored by identical repetition (SRT_Id) and correct target structure (SRT_Tar) as dependent variables to determine whether clinical group membership had a statistical effect. Both measures gave significant results SRT_Id: $[\chi^2(3, N = 106) = 67.6, p < .001, \eta^2 = .633]$,

SRT_Tar: [$\chi^2(3, N = 106) = 70.5, p < .001, \eta^2 = .65$]. Subsequently, pairwise comparisons were run using *post hoc* Mann-Whitney U tests. To reduce Type I error from multiple comparisons, the Bonferroni adjustment was applied to the obtained *p*-values. The groups' overall performance is depicted in **Figure 24**. On the whole task, monolingual and bilingual typically developing children performed significantly better than their peers with SLI. Results of Study 2 are replicated with the larger sample of this study. Both SRT measures sharply distinguish between MoTD and MoSLI: SRT_Id ($U = .000, p < .001, r = .847$), SRT_Tar ($U = .000, p < .001, r = .847$), BiTD and BiSLI SRT_Id ($U = 44.0, p < .001, r = .699$), SRT_Tar ($U = 27.5, p < .001, r = .727$) and most importantly between MoSLI and BiTD: SRT_Id ($U = 151.0, p < .001, r = .579$), SRT_Tar ($U = 112.5, p < .001, r = .639$). No significant differences were found between MoSLI and BiSLI. However, a significant difference with medium effect size was found between the MoTDs and BiTDs for both scoring methods, with a larger effect size for SRT_Id compared to SRT_Tar: SRT_Id ($U = 85.0, p < .001, r = 0.597$), SRT_Tar ($U = 118.5, p < .001, r = .532$).

Receiver Operating Characteristic (ROC) analyses were conducted on the monolingual and bilingual data separately (see **Table 25** and **Figure 25**) and revealed excellent diagnostic accuracy with 100% sensitivity and specificity for both SRT measures, “identical repetition” and “target structure”, for monolingual children. As can be seen in **Table 25**, the optimal cutoffs (i.e., scores associated with the highest diagnostic accuracy) for SRT_Id were lower than for SRT_Tar in both monolingual and bilingual groups and were generally much lower for the bilingual children (SRT_Id, 32.2%, and SRT_Tar, 52.2%) compared to the monolingual children (SRT_Id, 61.1%, and SRT_Tar, 75.5%). Diagnostic accuracy yielded good results for both measures, with excellent sensitivity (100%, i.e., all BiSLI children were identified as language impaired by the task) for both SRT_Id and SRT_Tar and fair specificity (SRT_Id, 83.7%, SRT_Tar, 88.4%). Given its better specificity, the overall diagnostic accuracy for SRT_Tar was better than SRT_Id in the case of bilingual children (91% for SRT_Id vs. 94.2% for SRT_Tar), confirming previous findings of studies using the German LITMUS-SRT that scoring by correct target structure is better suited for the assessment of bilingual children (Abed Ibrahim & Hamann, 2017; Abed Ibrahim et al., 2018; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018). Given its better diagnostic accuracy, subsequent analyses will henceforth be limited to SRT_Tar.

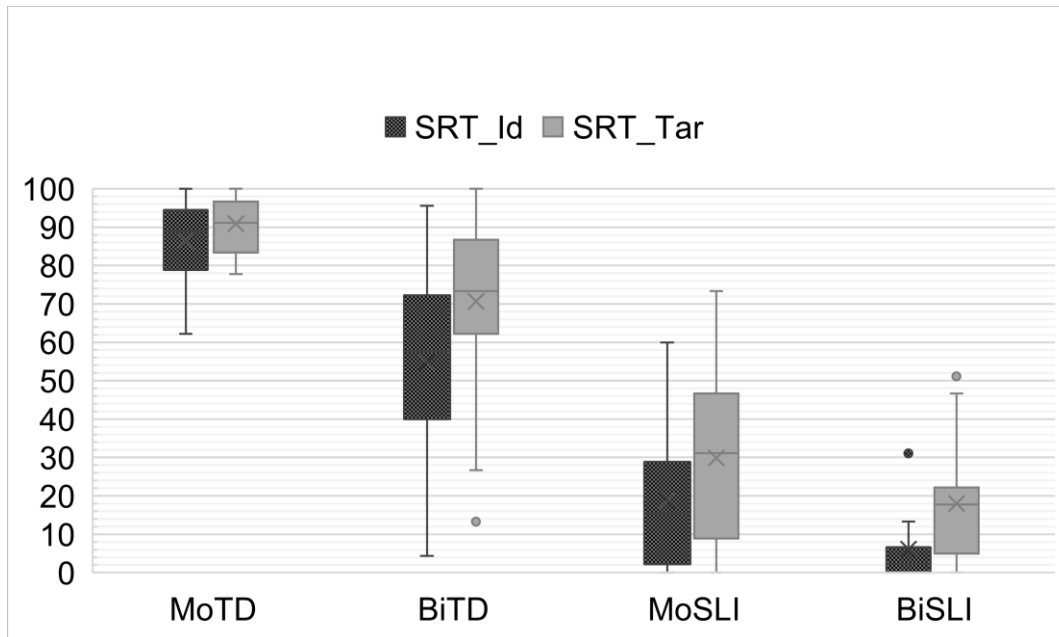


Figure 24: German LITMUS-SRT: Percentage identical repetition and correct target structure

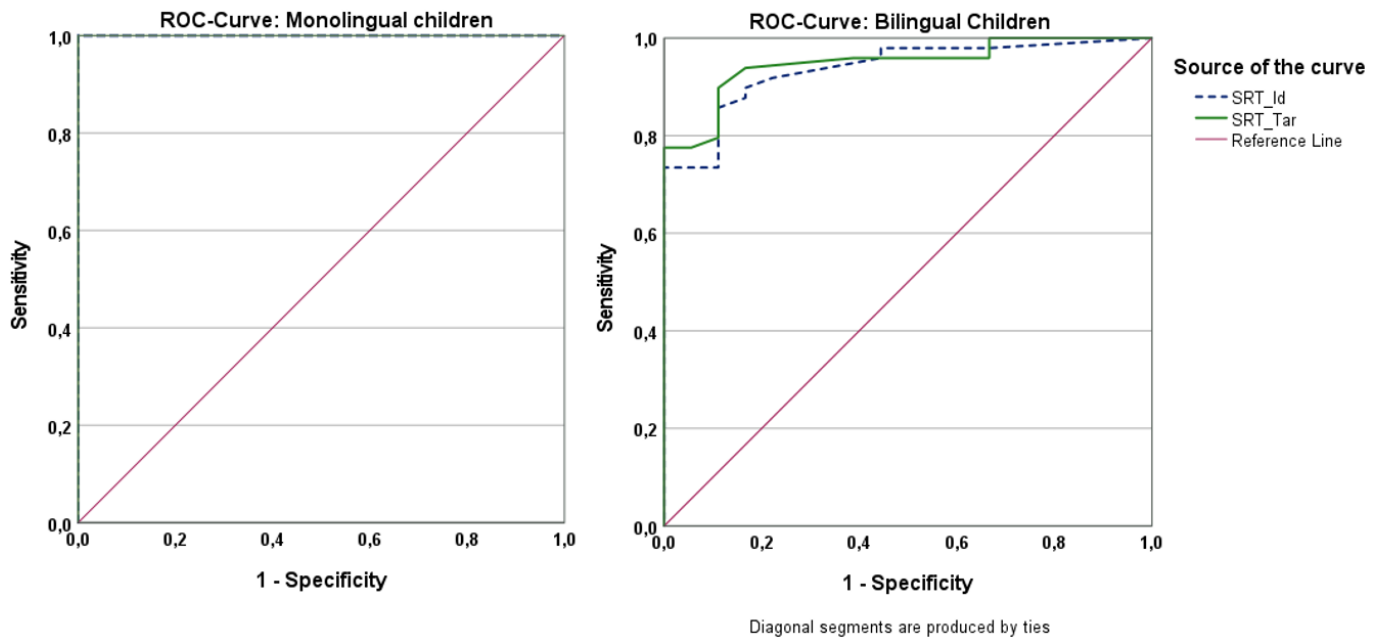


Figure 25: ROC curves for German LITMUS-SRT: Identical repetition and correct target structure for monolingual (left) and bilingual (right) children

Table 25: Diagnostic accuracy of the German LITMUS-SRT in monolingual and bilingual children: identical repetition (SRT_Id) vs. correct target structure (SRT_Tar)

	Monolinguals		Bilinguals	
	SRT_Id	SRT_Tar	SRT_Id	SRT_Tar
AUC	1.000	1.000	.937	.948
Cutoff (%)	61.1	75.5	32.2	52.2
Sensitivity (%)	100	100	100	100
Specificity (%)	100	100	83.7	88.4
LR+	undefined	undefined	6.13	8.62
LR–	0.0	0.0	0.0	0.0

Notes. AUC (Area Under Curve) values between .90 and 1.0 indicate excellent accuracy, values between .80 and .90 good diagnostic accuracy, and values between .70 and .80 fair diagnostic accuracy, whereas values below .60 indicate a worthless test (Swets et al., 2000).

^bLR+ values of 1-2 are indicative of a minimal increase in likelihood for a score below the cutoff to belong to an SLI case, 2-5 a small increase, 5-10 a moderate increase, and values above 10 indicate a large increase in likelihood (Dollaghan, 2004; Grimes & Schultz, 2005).

^cLR–values of 0.5-1.0 are indicative of a minimal increase in likelihood for a score above the cutoff to belong to a TD case, 0.2-0.5 a small increase, 0.1-0.2 a moderate increase, and a value below 0.1 indicates a large increase in likelihood (Dollaghan, 2004; Grimes & Schultz, 2005).

6.3.2 Effects of syntactic complexity

Figure 26 and **Table 26** compare the repetition accuracy (scored by correct target structure: SRT_Tar) of the structures featured in the German LITMUS-SRT, grouped by level of complexity. The results of pairwise comparisons are given in **Table 27**. As shown in **Figure 26** and **Tables 26-27**, accuracy scores decrease as syntactic complexity increases in all groups; however, the effect is much more pronounced in the MoSLI and BiSLI groups, which perform significantly less accurately than their MoTD and BiTD counterparts on all levels, showing almost floor performance on Level 2 and Level 3 structures. Importantly, there was no overlap between BiTDs and either MoSLI or BiSLI groups on any of the complexity levels.

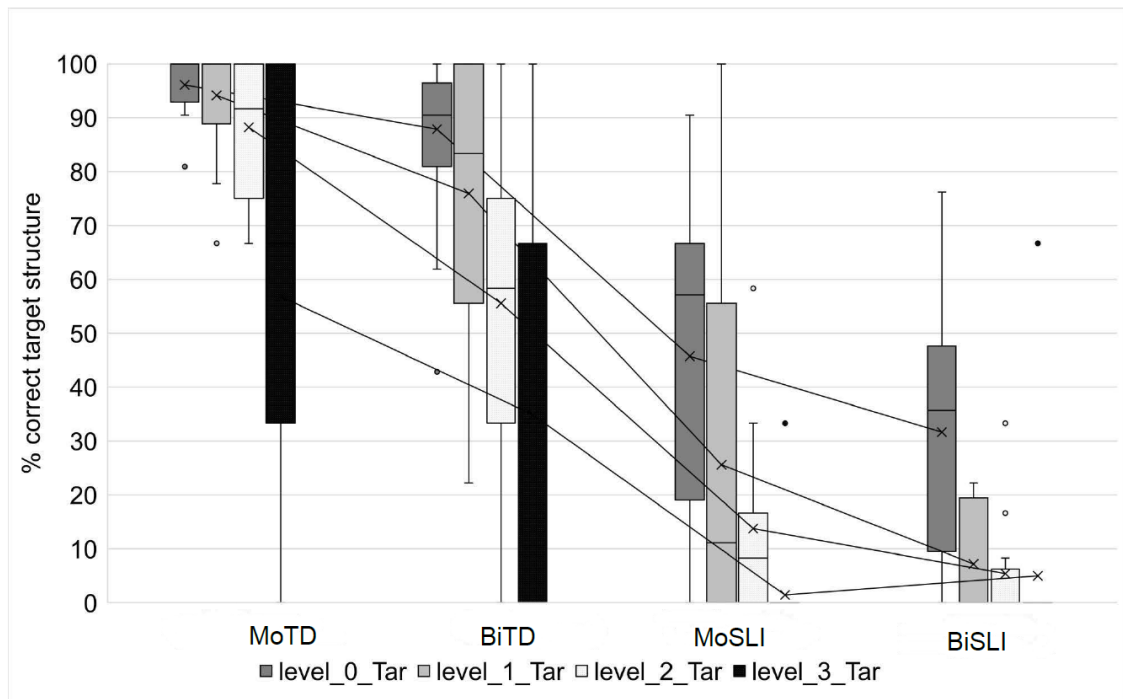


Figure 26: % SRT_Tar for sentences with and without computational complexity factors

Table 26: Mean (%) SRT_Tar and standard deviation by complexity level

	MoTD	BiTD	MoSLI	BiSLI
Level 0	96.1 (5.4)	86.9 (14.4)	45.8 (30.7)	30.8 (23.5)
Level 1	94.1 (9.7)	73.7 (26.5)	25.6 (30.8)	6.4 (9.3)
Level 2	88.2 (13.5)	54.3 (26.6)	13.8 (17.3)	5.7 (11.1)
Level 3	56.9 (38.7)	31.9 (37.8)	1.4 (7.0)	5.3 (16.7)

Although MoSLIs appear to have slightly higher means than their BiSLI peers on Levels 0, 1, and 2, no significant differences emerged between MoSLIs and BiSLIs for any of the complexity levels. However, significant differences with a small effect size for Level 1 and a medium to large effect size for Level 2 emerged between MoTDs and BiTDs (see **Table 27**). Level 3, which consists only of the most complex construction, namely object relatives with intervention (embedding, *wh*-movement, and intervention), was difficult for all groups. Even monolingual TDs scored only slightly above 50 percent, and a significant proportion of the bilinguals performed below chance. Monolingual and bilingual children with SLI, however, showed floor performance.

Table 27: The effect of (clinical) group membership on different complexity levels of the SRT and post hoc pairwise comparisons

	Level-0	Level-1	Level-2	Level-3
	$\chi^2(3, N=106)$ = 64.708 $p < .001$	$\chi^2(3, N=106)$ = 65.189 $p < .001$	$\chi^2(3, N=106)$ = 68.452 $p < .001$	$\chi^2(3, N=106)$ = 35.141 $p < .001$
MoTD vs. MoSLI	$U = 6.50$ $p < .001$ $r = .824$	$U = 16.0$ $p < .001$ $r = .792$	$U = 0.00$ $p < .001$ $r = .856$	$U = 38.5$ $p < .001$ $r = .784$
MoTD vs. BiSLI	$U = 0.00$ $p < .001$ $r = .861$	$U = 0.00$ $p < .001$ $r = .883$	$U = 0.00$ $p < .001$ $r = .884$	$U = 0.00$ $p < .001$ $r = .706$
BiTD vs. MoSLI	$U = 97.5$ $p < .001$ $r = .661$	$U = 127.0$ $p < .001$ $r = .618$	$U = 104.0$ $p < .001$ $r = .650$	$U = 270.0$ $p < .001$ $r = .463$
BiTD vs. BiSLI	$U = 19.5$ $p < .001$ $r = .758$	$U = 6.00$ $p < .001$ $r = .787$	$U = 43.0$ $p < .001$ $r = .716$	$U = 251.0$ $p < .001$ $r = .378$
MoTD vs. BiTD	$n.s.$	$U = 205.0$ $p < .05$ $r = .363$	$U = 100.0$ $p < .001$ $r = .566$	$n.s.$
MoSLI vs. BiSLI	$n.s.$	$n.s.$	$n.s.$	$n.s.$

Given that Level 3 featured only one sentence type and was problematic for all groups, within-group analyses were restricted to Levels 0-2. A Friedman test revealed an overall significant effect of complexity level in all groups: MoTD [$\chi^2(2) 2.78, p < .001$], BiTD [$\chi^2(2) 43.532, p < .001$], MoSLI [$\chi^2(2) 25.787, p < .001$], BiSLI [$\chi^2(2) 23.310, p < .001$]. Post hoc pairwise comparisons using the Wilcoxon test with Bonferroni Holm correction showed significant differences for the BiTD group between Level 0 and Level 1 ($Z = -3.862, p < .001, \eta^2 = .346$), Level 0 and Level 2 ($Z = -5.472, p < .001, \eta^2 = .696$) and Level 1 and Level 2 ($Z = -4.513, p < .001, \eta^2 = .440$). The effect size was larger when two complexity factors were present than when only one was present. The MoTD group, on the other hand, showed only a marginally significant effect for Level 0 versus Level 2 ($Z = -2.314, p = .063, \eta^2 = .232$), while the other comparisons were not significant. As to the SLI groups, both groups showed a pattern similar to the BiTDs: a significant difference emerged between Level 0 and Level 1 and between Level 0 and Level 2, but not between Level 1 and Level 2; MoSLI: Level 0 versus Level 1 ($Z = -3.213, p < .001, \eta^2 = .448$), Level 0 versus Level 2 ($Z = -3.921, p < .001, \eta^2 = .668$); BiSLI: Level 0 versus Level 1 ($Z = -3.293, p < .001, \eta^2 = .570$), Level 0 versus Level 2 ($Z = -3.367, p < .001, \eta^2 = .570$).

= .596). Again, the effect size was larger when two complexity factors were present than when only one was present.

Since a significant between-group difference emerged between MoTDs and BiTDs for Levels 1 and 2, a structure-by-structure analysis was conducted to identify which sentence types caused the significant differences between MoTDs and BiTDs. For Level 1, significant differences with relatively small effect sizes emerged between MoTDs and BiTDs for long passives ($U = 235.0, p < .05, r = .334$) and finite complement clauses ($U = 235.0, p < .05, r = .327$), but not for object who-questions. For Level 2, the MoTD and BiTD groups differed significantly on all sentence types: object which-questions: ($U = 206.5, p < .05, r = .380$), subject relatives: ($U = 184.0, p < .05, r = .431$), object relatives without intervener: ($U = 188.0, p < .05, r = .397$), topicalizations: ($U = 149.0, p < .001, r = .481$). The performance of all groups¹⁷⁵ on the structures in Levels 1 and 2 is illustrated in **Figures 27** and **28**. Note that **Figure 28** includes object relatives with intervention.

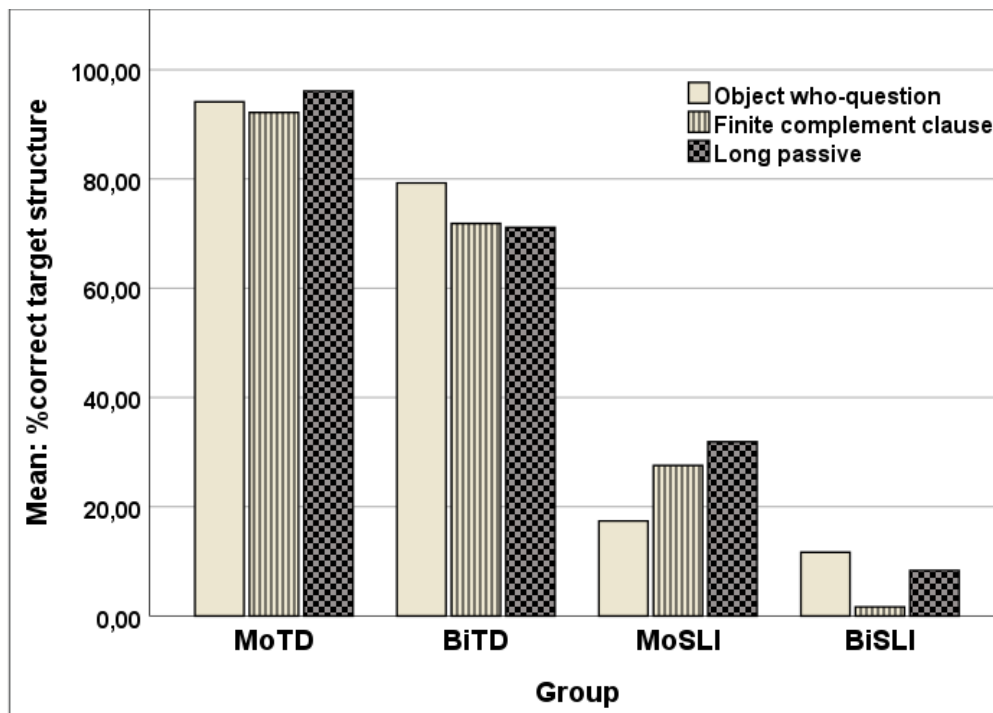


Figure 27: Complexity Level 1: Percentage of sentences with correct target structure

¹⁷⁵ Given that SLI children performed poorly on all structures, the pairwise comparisons are restricted to the MoTD and BiTD groups.

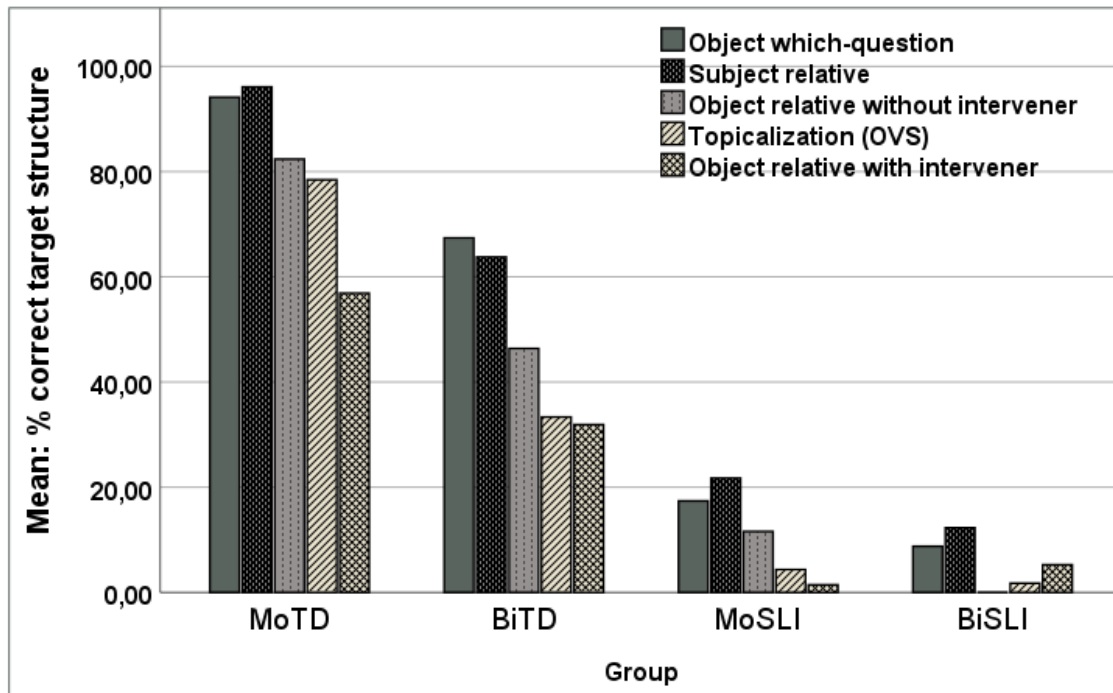


Figure 28: Complexity Levels 2 & 3: Percentage of sentences with correct target structure

Since Hamann et al.'s (2017) study found older BiTDs to perform on par with MoTDs, it was speculated that age affects performance. Hence, we split the BiTD group by age into two groups: a younger BiTD group (age range: 5;6-6;11), which comprises preschool children and first graders, and an older BiTD group (age range: 7;0-8;11), which only includes school-age children. Interestingly, significant differences relative to MoTDs emerge only for the younger BiTD group on all structures of Level 2 with large effect sizes for ORs and topicalizations: object which-questions: ($U = 77.0, p < .05, r = .491$), subject relatives: ($U = 80.0, p < .05, r = .490$), object relatives without intervener: ($U = 51.5, p < .001, r = .596$), topicalizations: ($U = 33.0, p < .001, r = .706$). With respect to Level 1, significant differences between MoTDs and the younger BiTD group were found for finite complement clauses ($U = 78.0, p < .05, r = .489$) and long passives ($U = 84.0, p < .05, r = .472$), see **Figures 29 and 30**.

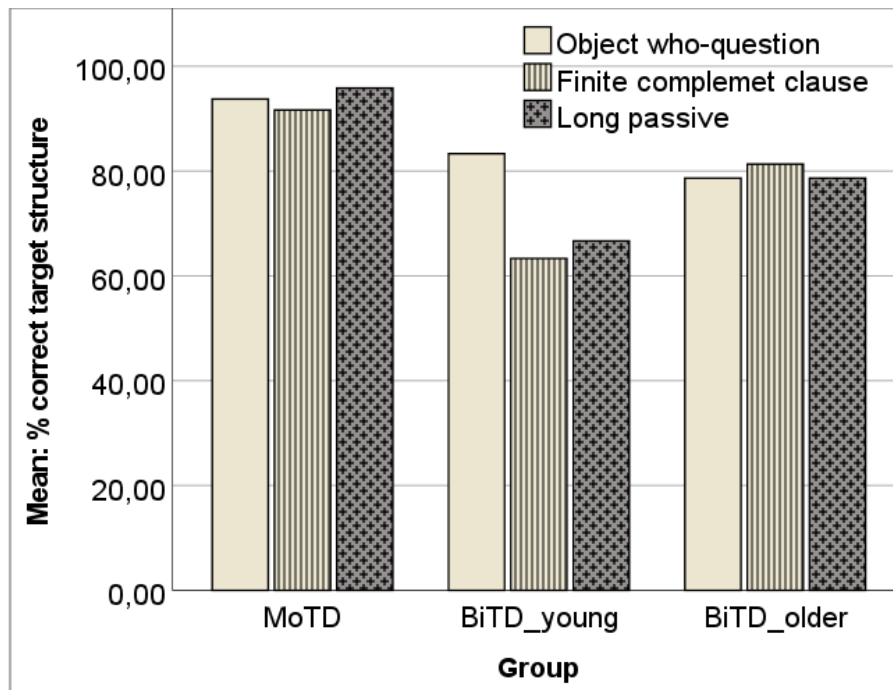


Figure 29: Complexity Level 1: Percentage of sentences with correct target structure (BiTDs split by age)

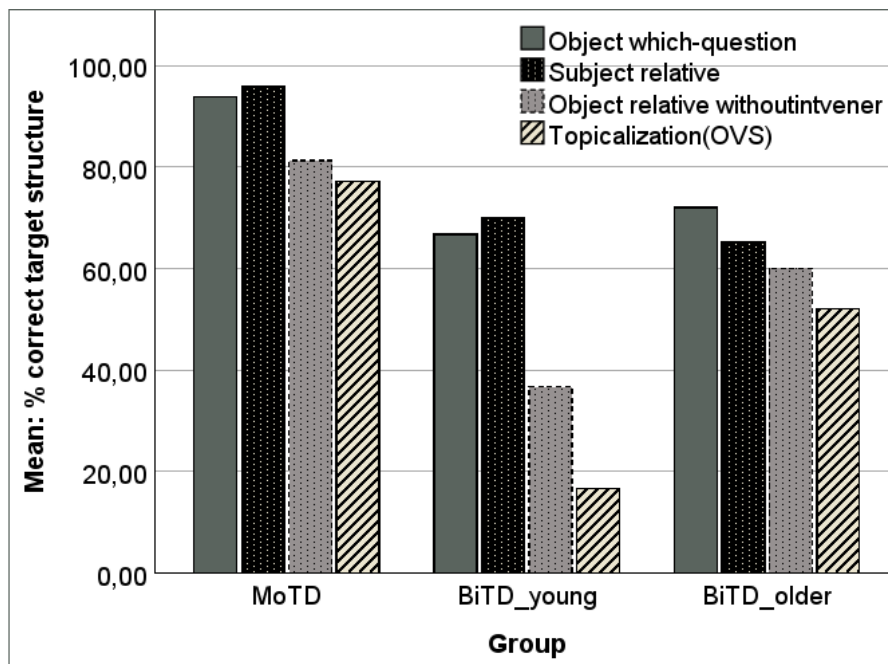


Figure 30: Complexity Level 2: Percentage of sentences with correct target structure (BiTDs split by age)

6.3.3 Factors predicting performance on German LITMUS-SRT

The third research question asked to what extent the performance of monolingual and bilingual children with and without SLI on the LITMUS-SRT was associated with linguistic knowledge and/or WM capacity. Given the observed interaction between age and performance on complex structures in the BiTD group, partial Spearman correlations controlling for chronological age were run (for each group separately) to check for correlations between SRT and language and WM measures. The language measures considered for the analyses were expressive and receptive vocabulary (WWT, Glück, 2011) and the children's developmental level in the LiSe-DaZ subscale assessing sentence complexity in elicited production (Schulz & Tracy, 2011). As to WM measures, FDS and BDS (indexed by the highest span attained by the child) were included as measures of VSTM and VWM, respectively. Variables yielding significant bivariate correlations with the SRT measures were considered for subsequent hierarchical regression analyses and were entered into the blocks of the model in the order of strength of correlation. To avoid multicollinearity between L2 expressive and receptive vocabulary, both factors were treated in separate models (Dormann et al., 2013). Due to the very poor performance and lack of variance of MoSLIs and BiSLIs on the structures featured in Levels 1 and especially Level 2, analyses on the SLI groups are only restricted to the global measure SRT. Note that, though not reported here, the performance of the MoSLIs and BiSLIs on the individual complexity levels showed no significant correlations with WM measures; hence, WM could not have emerged as a significant predictor in a regression model.

As to the MoTDs, a single strong positive correlation emerged between receptive vocabulary ($r = .552, p < .05$) and overall SRT performance, while FDS showed only a marginally significant correlation with SRT ($r = .512, p = .051$) and was therefore not included in the regression analysis. Hierarchical regression analyses showed that 34.2% of the variance of MoTDs' performance on the SRT is explained by receptive vocabulary ($\beta = 0.585, T = 2.695, p < .05$). Likewise, Level 1 was strongly correlated only with receptive vocabulary, which accounted for 52% of the variance in performance ($\beta = 0.721, T = 3.895, p < .01$). In contrast, the structurally more complex Level 2 was only correlated with FDS ($r = .619, p < .05$), explaining 49.4% of the variance ($\beta = 0.703, T$

=3.829, $p < .01$). Similar to their MoTD peers, the performance of the BiTDs on the SRT showed medium positive correlations with expressive vocabulary ($r = .414, p < .01$) and receptive vocabulary ($r = .471, p < .01$). However, unlike their MoTD peers, a strong positive correlation emerged for FDS ($r = .629, p < .001$). Results of the regression analyses showed that a substantial proportion of the variance in performance on SRT was explained by FDS, which accounted for 44.3% of the variance. Adding expressive and receptive vocabulary in Step 2 explained an additional 6.3% and 5.3% of the variance, respectively (see **Table 28**).

Table 28: Summary of hierarchical regression analysis for linguistic and WM factors predicting the performance of BiTD children on the SRT

		Model 1				Model 2		
		<i>b</i>	<i>SE b</i>	β		<i>b</i>	<i>SE b</i>	β
Step 1								
	Constant	1.812	11.960		Constant	2.280	12.147	
	FDS	15.439	2.639	.666***	FDS	15.363	2.674	.663***
Step 2	Constant	15.405	12.829		Constant	24.347	15.839	
	FDS	13.727	2.623	.592***	FDS	11.687	3.130	.505**
	Expressive vocabulary	4.194	1.816	.261*	Receptive vocabulary	2.794	1.352	.279*

Model 1: $R^2 = .443$ for Step 1: $\Delta R^2 = .063$ for Step 2 ($p = .026$)

Model 2: $R^2 = .443$ for Step 1: $\Delta R^2 = .053$ for Step 2 ($p = .045$)

* $p < .05$. ** $p < .01$. *** $p < .001$.

A similar correlational pattern as for global performance was observed for BiTDs for Levels 1 and Level 2, which were also positively correlated with FDS and expressive and receptive vocabulary: Level 1: expressive vocabulary ($r = .417, p < .01$), receptive vocabulary ($r = .526, p < .001$), FDS ($r = .504, p < .01$), Level 2: expressive vocabulary ($r = .416, p < .01$), receptive vocabulary ($r = .326, p < .05$), FDS ($r = .595, p < .001$). Interestingly, Level 2 was more highly correlated with FDS than with vocabulary measures. Regression analysis of Level 1 (**Table 29**) showed that FDS explained 25.9% of the variance in Step 1, with expressive vocabulary explaining an additional 12.2% in Step 2. Although FDS explains 25.4% of the variance in the first step of the second model, it becomes insignificant once receptive vocabulary is entered in Step 2, explaining further 14.2% of the variance. The structurally more complex Level 2 shows a different pattern, where performance is only predicted by FDS, which accounts for 34% of the variance

with no further contributions for receptive or expressive vocabulary ($\beta = 0.583$, $T = 4.487$, $p < .001$).

Table 29: Summary of hierarchical regression analysis for linguistic and WM factors predicting the performance of BiTD children on Level 1 structures of the German LITMUS SRT

		Model 1				Model 2		
		<i>b</i>	<i>SE b</i>	β		<i>b</i>	<i>SE b</i>	β
Step 1								
	Constant	8.878	17.797		Constant	10.012	18.067	
	FDS	14.717	3.892	.508***	FDS	14.530	3.940	.504**
Step 2	Constant	32.230	18.438		Constant	53.408	21.841	
	FDS	11.741	3.753	.406**	FDS	7.345	4.306	.255 $p = .096$
	Expressive vocabulary	7.066	2.512	.365**	Receptive vocabulary	5.573	1.842	.451**

Model 1: $R^2 = .259$ for Step 1: $\Delta R^2 = .122$ for Step 2 ($p = .008$)

Model 2: $R^2 = .254$ for Step 1: $\Delta R^2 = .142$ for Step 2 ($p = .004$)

* $p < .05$. ** $p < .01$. *** $p < .001$.

As to the MoSLI children, strong positive correlations were found between SRT and all language measures: expressive vocabulary ($r = .591$, $p < .01$), receptive vocabulary ($r = .686$, $p < .01$), and developmental level ($r = .685$, $p < .01$). As to WM, no significant correlations were found for either WM measure. As can be seen in the model summaries in **Table 30**, L2 morphosyntax (indexed by the developmental level on the LiSe-DaZ, see **3.2.1.2**) emerged as the sole factor explaining most of the variance in the performance of MoSLIs. Although receptive vocabulary explains 27.5% of the variance when entered in Step 1 of Model 1, it became insignificant once L2 morphosyntax is entered in Step 2, accounting for 35% of the variance. Expressive vocabulary did not have a significant contribution either, with L2 morphosyntax explaining 59.7% of the variance in Model 2. With respect to the BiSLI group, performance on the SRT was only positively correlated to L2 morphosyntax ($r = .552$, $p < .05$), which explained 56.7% of the variance in performance ($\beta = 0.753$, $T = 5.242$, $p < .001$).

Table 30: Summary of hierarchical regression analysis for linguistic and WM factors predicting the performance of MoSLI children on German LITMUS SRT

		Model 1			Model 2			
		<i>b</i>	<i>SE b</i>	β		<i>b</i>	<i>SE b</i>	β
Step 1								
	Constant	45.277	7.168		Constant	-75.257	19.528	
	Receptive vocabulary	9.623	3.495	.524*	Developmental level	30.352	5.574	.773***
Step 2	Constant	-56.779	24.795		Constant	-53.058	22.367	
	Receptive vocabulary	3.519	2.957	.192 <i>p</i> = .249	Developmental level	26.577	5.706	.677***
	Developmental level	26.663	6.329	.679***	Expressive vocabulary	5.874	3.306	.258 <i>P</i> = .092

Model 1: $R^2 = .275$ for Step 1: $\Delta R^2 = .350$ for Step 2 ($p = .012$)

Model 2: $R^2 = .597$ for Step 1: $\Delta R^2 = .057$ for Step 2 ($p = .092$)

* $p < .05$, ** $p < .01$, *** $p < .001$.

6.4 Discussion

Diagnostic accuracy

The study's first aim was to verify whether the good diagnostic accuracy previously reported for the German LITMUS-SRT, particularly when scored by correct target structure, can be replicated with a larger participant sample and whether comparable cutoffs can still be obtained. For this purpose, Receiver Operating Characteristic curve (ROC) analysis was employed to calculate sensitivity and specificity levels for monolinguals and bilinguals separately based on the children's clinical status as verified by the comprehensive procedure described in 3.1.1. The finding that SRT scored by correct target structure is better suited than SRT scored by identical repetition for the assessment of language abilities of bilingual children with German as L2 was replicated in this study, given its better specificity and overall diagnostic accuracy than the former scoring measure. Importantly, SRT shows 100% sensitivity for SLI for both monolingual and bilingual children, showing that the task is well constructed and can identify SLI regardless of the child's lingual status. Moreover, the cutoff score for SRT_Tar, which is associated with the highest diagnostic accuracy (52.2%), resembles the cutoff obtained in previous studies featuring a subset of the current participant sample (Chilla et al., 2021; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018) and is very close to the cutoff obtained via unsupervised machine learning in Study 2. This is a welcome result as it

validates the robustness of the task. Note that, similar to many other studies featuring LITMUS sentence repetition tools (e.g., Armon-Lotem & Meir, 2016; Fleckstein et al., 2018), the cutoff scores separating TD from SLI in bilinguals are lower than those for monolinguals, implying that monolingual norms do not apply to bilinguals even in carefully constructed tests like LITMUS-SRTs. The reduced specificity of the task was explicitly addressed by Hamann and Abed Ibrahim (2017), who identified language dominance as a contributing factor. This factor is also speculated to have contributed to the reduced specificity in the current sample, given that 24 out of 46 BiTDs are L1-dominant according to PaBiQ estimates. Although MoTDs outperformed BiTDs, the task achieved a diagnostic accuracy of 94.2% and, importantly, there was no overlap between BiTDs and either SLI group, which is promising from a clinical perspective.

Computational complexity and WM effects

The second and third research questions addressed the effects of syntactic complexity on the performance of monolingual and bilingual children with and without language impairment. For this purpose, the sentences of the German LITMUS-SRT were grouped into four different complexity levels, where the complexity of a given sentence type was estimated by the number of morphosyntactic operations involved in its derivation, in addition to the presence of intervention (NP-restriction) as an additional complexity factor. In these comparisons, the measure of “correct target structure” was used, which ignores lexical substitution errors and systematic case and gender errors that do not interfere with the realization of the targeted construction.

As predicted, the repetition accuracy decreased in all groups as a function of cumulative syntactic complexity (**Figure 26 & Table 26**). This is in line with previous studies that show that recall accuracy decreases as processing load increases due to the presence of complexity factors (e.g., Friedmann & Reznick, 2021). Although MoSLI and BiSLI children performed poorly on all complexity levels, they experienced more difficulties with structures of Level 1 (object who-questions, finite complement clauses, and long passives), which involved one complexity factor and could not cope with the presence of two or more complexity factors (object which-questions, subject relatives, object relatives without and with intervener and topicalizations) as indicated by their floor performance on Level 2 and Level 3. It is important to note that Level 3 featured only object relatives with intervention, which were significantly more difficult for all groups.

Even the MoTDs performed only slightly above chance, and the BiTDs slightly below chance, with no significant differences between them. This is in line with Ruigendijk and Friedmann (2017), who reported difficulties with object relatives across different tasks, including sentence repetition, until age 11. Children with SLI, however, perform at 0%. Using a subset of the data, Abed Ibrahim and Hamann (2017) analyzed frequent error patterns on ORs with intervention and found the primary source for errors in MoTDs and BiTDs to be erroneous case marking on the relative head pronoun, e.g., double accusative marking or resorting to canonical word order by producing subject relatives instead of object relatives, hence reversing the thematic roles. In contrast, children with SLI suffered a total computational breakdown, producing sentential fragments or monoclausals. This also lends support to the analysis of the exceptional difficulty children encounter with intervention, as suggested by Friedmann et al. (2009) and by Belletti's (2014) and Belletti and Rizzi's (2012) explanation for the children's most frequent compensatory strategy, namely, avoiding intervention by producing passives (smuggling) instead of ORs in elicitation tasks. In sentence repetition, ORs cannot be avoided; thus, when they are produced, commission errors such as those previously mentioned occur (see Hamann et al., 2017, for a discussion).

The significant difference that was observed between BiTDs and MoTDs on the SRT task as a whole turned out to be the result of significant differences between BiTDs and MoTDs on Level 1 and Level 2 (see **Table 27**). Analyses of the individual structures of the Levels revealed a significant difference for long passives (Level 1), whereas object-who questions and finite complement clauses showed no significant differences. However, MoTDs significantly outperformed BiTDs on all constructions featured in Level 2 (object who-questions, subject relatives, object relatives without intervention, and topicalizations), with the largest effect size for topicalizations. Interestingly, when the BiTD children were split into two age groups, significant differences to MoTDs emerged only for BiTDs younger than 7, whereas older BiTDs performed on par with their MoTD peers (see **Figures 29 & 30**). This finding aligns with Hamann et al.'s (2017) study, which showed that younger BiTDs still encountered problems with SRs, ORs, and long passives, whereas older bilingual children appear to catch up with their monolingual peers. These findings also corroborate Biran and Ruigendijk (2015), who found that monolingual children had difficulties with topicalizations until age seven, and align with Ruigendijk

and Friedmann (2017), who showed that subject relatives and passives are comprehended well by monolingual children from age seven onward.

All the structures of Level 2 (except for SRs) involve A'-movement, which means that theta-role assignment cannot follow from (canonical) word order. Since the latter sentences did not incorporate a mismatch in the number features, theta roles had to be assigned by using case markings on the subject and object DPs as disambiguating cues (Stegenwallner-Schütz & Adani, 2021). Although the younger BiTDs did not show mastery of the latter constructions, they still significantly outperformed the SLI groups, who performed very poorly on all of them. Moreover, the larger effect sizes observed for the presence of two complexity factors instead of only one in all groups, with still larger effect sizes in the SLI groups, confirm that the cumulation of syntactic operations is more taxing for processing resources. Notably, no significant within-group effects emerged for the comparison 'Level 1 versus. Level 2' for either SLI group, whereas a significant effect emerged for the BiTDs. For the MoSLI and BiSLI groups, significant within-group effects only emerged when Level 0 was contrasted with Level 1 or Level 2, but not when Level 1 was compared to Level 2. This implies that the presence of one complexity factor suffices to cause a computational breakdown, so that adding a further factor does not result in significant within-group effects in the case of children with SLI. On the other hand, the lack of a within-group effect in the MoTD group can be explained by their homogeneous and good performance on both Level 1 and Level 2.

The second aim was to explore whether monolingual and bilingual children with and without SLI demonstrate different patterns of association between overall performance on SRT and on the different complexity levels with other linguistic measures and VSTM and VWM measures. In this respect, Zebib et al.'s (2019) finding concerning the unique role of the complex span task, i.e., VWM as a predictor for performance on French LITMUS-SRT, could not be replicated in this study, as neither WM measure yielded significant correlations with any of the SRT measures. Interestingly, it was L2 morphosyntax, namely the syntactic developmental level from the LiSe-DaZ, which predicted global performance of the BiSLI and MoSLI groups. This result corroborates Hamann et al.'s (2020) results, showing that a developmental level of 4 (ability to produce embedded constructions) is necessary for performance above the pathology cut-off on the German LITMUS-SRT. However, the lack of significant correlations between SRT and WM does not mean that WM is not implicated in sentence repetition. A close inspection

of the within-group distribution of FDS and BDS scores of the children with SLI investigated here revealed that the majority had a span of 3 on FDS and almost floor performance on BDS, which reduced the inter-individual variability and made finding an association with performance difficult. As to the TD groups, performance of the MoTDs was explained by receptive vocabulary, which accounted for 34.2% of the variance on the task as a whole and 52% of the variance in Level 1. However, when complexity increased, MoTD children seemed to rely more on their VSTM abilities, which explained 49.2% of the variance on Level 2, confirming that WM is strongly implicated in processing computationally more complex constructions. As to the BiTDs, a combination of expressive/receptive vocabulary explained global performance, where FDS (VSTM) contributed most towards predicting the variance, explaining 44.3% of it. A similar pattern emerged for Level 1, where a combination of expressive/receptive vocabulary and FDS explained 38.1-39.6% of the variance. Interestingly, performance on Level 2 structures was only accounted for by FDS, confirming that processing computationally more complex structures necessitates more WM capacities. The absence of an effect for L2 morphosyntax in MoTDs and BiTDs can be explained by the lack of variance, as all TD children had reached level 4. The fact that BiTDs rely more on their WM resources to process complex constructions than MoTDs may result from their weaker L2 vocabulary skills, which are a common area of weakness for bilinguals when single-language vocabulary is considered. While MoTDs can rely on their more advanced language skills (i.e., language knowledge in long-term memory), BiTD children compensate for their weaker L2 abilities by resorting to their intact WM capacities (see Peets & Bialystok, 2010).

6.5 Conclusion

The findings of this study have confirmed that aspects of computational complexity are particularly challenging for children with SLI, whose weak linguistic representations and impaired WM capacities are overtaxed by them. BiTD children significantly outperformed MoSLIs and BiSLIs with no overlap, confirming the prediction that “typical bilinguals have an intact language faculty and working memory and can deal with movement, embedding, and recursion” (Hamann et al., 2017, p. 28). Bilingual

children use complex constructions, and as demonstrated by the performance of the older BiTD group, need significantly less exposure to the L2 to pattern like their monolingual age-matched peers compared to other input-sensitive measures like vocabulary and inflectional morphology. What clearly emerges from the study is that using complex structures involving movement, embedding, and intervention can sharply distinguish between typical and impaired language development in children, regardless of their lingual status. Hence, tasks incorporating such complex constructions offer promising tools for establishing typical language development in bilingual children.

However, it also turns out that using the most complex constructions, such as ORs with intervention, does not always discriminate between SLI and TD, since many TD children, at least in the age range investigated here, also show floor performance. In contrast, performance on the SRT task as a whole, which also features simpler constructions, achieved the desired result since there was no overlap between BiTD and MoSLI or BiSLI. These results are of great relevance for the development of shorter German LITMUS_SRT tasks, which, in the interest of easier administration of such tasks by SLPs, is one of the next aims of this research program.

7 General discussion

7.1 Summary

As outlined in Chapter 1, this dissertation was motivated by the paucity of appropriate language assessment tools for SLI in bilingual populations. This is particularly relevant for the German context (Chilla, 2019; Chilla & Hamann, 2021; Deml et al., 2018; Hertel et al., 2022), as it plays an important role across all stages and sectors of education. Not only does language assessment determine whether a child is eligible for special needs services in the form of speech-language therapy, but it also informs decisions concerning the transition from kindergarten to elementary school and other forms of schools in the case of older children. Moreover, reliable assessment is important because children with language impairment need different interventions than BiTD children whose intrinsic language capacity is intact (Roy et al., 2014). It also plays a central role in guiding therapy and tailoring intervention choices to the child's needs (De Lamo White & Jin, 2011).

Chapter 2 highlighted the diagnostic challenges in bilingual contexts, which on the one hand are attributable to the substantial heterogeneity of children with SLI, who present diverse linguistic profiles and deficits of varying severity across language domains (Conti-Ramsden et al., 2001a; Leonard, 2014a), and on the other hand to the great variability of typical bilingual language development, which is influenced by the complex interplay of child internal and external factors (Hamann, 2012; Paradis, 2023). A major source for diagnostic problems is the general lack of suitable language assessment tools for bilingual children (Altman et al., 2022; Chilla & Hamann, 2021). Bilinguals are typically assessed with monolingual norm-referenced tests or informal measures that do not account for variation in bilingual experience and, with very few exceptions, lack bilingual norms (Chilla & Hamann, 2018; Voet Cornelli et al., 2013). Consequently, they often score below the monolingual typically developing average, which is misinterpreted as language impairment, resulting in overdiagnosis with SLI (Grimm & Schulz, 2014). The problem is further aggravated by the temporary overlap between clinical markers used to identify SLI in monolingual children and second language L2 interlanguage phenomena, making these markers unsuitable for diagnosing SLI in bilingual contexts (Chilla, 2008; Clahsen, 1991a; Hamann et al., 1998, 2017;

Rothweiler, 2006; Rothweiler et al., 2012). The key question is thus how to determine whether a bilingual child's below-average performance in the L2 reflects an underlying language impairment or is due to insufficient exposure to the societal language. In short, which assessment tools should be employed to avoid misdiagnosis and to minimize overlap between bilingual children with and without SLI? This dissertation comprises three empirical studies that address the latter question by evaluating recent recommendations and newly developed LITMUS assessment tools for resolving diagnostic challenges in multilingual contexts. The diagnostic accuracy of a given test or clinical marker is conventionally estimated by comparison with the best available reference (gold) standard used to diagnose the condition (Bossuyt et al., 2015; Umemneku Chikere et al., 2021). A particular strength of the studies reported in this thesis is that the clinical utility of the newly developed LITMUS tools was evaluated against a sample of monolingual and bilingual children with a rigorous clinical diagnosis that considered multiple language domains in both languages and accounted for bilingual exposure patterns. This approach distinguishes the present work from studies that rely solely on clinical referral or on standardized L2 assessment to establish clinical status (e.g., Boerma et al., 2015; Grimm, 2022; Wojtecka, 2020). Moreover, the number of first languages was limited to three widely spoken languages, reducing sources of cross-linguistic influence, a factor further verified in related BiLaD project studies (e.g., Abed Ibrahim et al., 2018; Chilla et al., 2021) within which this research arose.

The first study (Chapter 4) examined the recommendation to assess the heritage language (HL/L1) alongside the societal language (L2) to ensure a reliable diagnosis of SLI. Since a genuine language impairment affects language acquisition as a whole and should manifest in any language being acquired (Kohnert, 2010; Paradis, 2010a), it is recommended to assess a bilingual child in both languages when possible (ASHA, 2023; IALP, 2011; RCSLT, 2007), or at least in the child's dominant language (Fredman, 2006), which often corresponds to the heritage language (HL/L1) especially in early stages of bilingual language acquisition. However, bilingual children typically acquire their heritage language in a minority context under adverse input conditions with limited or no access to formal HL education. Consequently, they often experience a decline in heritage abilities after shifting to the societal language during school years and develop very limited or no HL literacy skills (Montrul, 2008, 2016a, 2016b). Additionally, HL interlanguage features can overlap with clinical markers of SLI in monolingual contexts

(Blom et al., 2019, 2022). Hence, the question arises of whether and how HL assessment can contribute to ruling in or out the language impairment in heritage bilingual children for whom the HL is not, or no longer, the dominant language.

Since assessment in the HL is often not feasible in clinical practice, clinicians and diagnostic professionals are forced to rely on L2 assessment; however, as shown in Chapter 2, appropriate norm-referenced tools for bilingual children are still lacking. In a cross-linguistic collaborative endeavor, a comprehensive set of cross-linguistic tools for identifying SLI in bilinguals was developed during COST Action IS0804. These tools, known as the LITMUS tools (Language Impairment Testing in Multilingual Settings, Armon-Lotem et al., 2015a), were constructed according to a set of (linguistic) principles that allow minimizing the effects of bilingualism while being highly indicative of the presence of SLI, and are available in more than 21 languages, including German. Two prominent tasks within the LITMUS toolbox are sentence repetition tasks (SRTs) and non-word repetition tasks (NWRs), complemented by the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015).

Depending on the construction, SRTs and NWRs assess not only verbal short-term memory (Archibald & Gathercole, 2006) but also, crucially, the command of syntactic and phonological representations/derivations (see Polišenská et al., 2015, for sentence repetition; Gallon et al., 2007, for nonword repetition), which are core domains affected in SLI. Compared to other types of language measures, sentence and nonword repetition are less constrained by pragmatic and discourse factors (Marinis & Armon-Lotem, 2017; Polišenská et al., 2015) and are widely used in clinical assessment. Hence, sentence and nonword repetition were selected as measures for cross-linguistic development within and beyond COST-Action IS0804 (Chiat, 2015; Marinis & Armon-Lotem, 2015), and LITMUS sentence repetition and nonword repetition tasks were specifically developed for identifying SLI in bilingual children acquiring German as a societal language (Hamann & Abed Ibrahim, 2017; Hamann & Chilla, 2018). The second study of this thesis (Chapter 5) investigated the diagnostic potential of the German versions of the repetition tasks and their robustness to the effects of language dominance for identifying SLI in simultaneous as well as in (early) successive preschool and primary-school-age bilingual children acquiring German as a societal language, which is not necessarily their dominant language at the time of assessment. Additionally, the study examined whether using a combination of the two tools and their sub-measures, each

focusing on a different aspect of language ability, enhances diagnostic accuracy for language impairment in bilingual settings.

Building on Study 2's findings, the third study examines how the specific construction of the German LITMUS-SRT contributed to the task's good diagnostic accuracy and robustness to language dominance, particularly when scored using the more lenient measure that assesses preservation of the target structure. As elaborated in **2.3.1.1**, the construction of the German LITMUS-SRT (see **2.1.2.2**) was guided by the notion of syntactic complexity: computationally complex constructions shown to be cross-linguistically vulnerable in children with SLI were incorporated into the task in addition to a set of language-specific constructions. Given their robust acquisition in typically developing bilinguals, these structures are expected to identify SLI in bilingual children with minimal overlap with MoSLI and BiSLI children, who, given their impaired language faculty and limited working memory capacity, are expected to struggle disproportionately with computationally complex constructions (Friedmann & Ruigendijk, 2017). Hence, Study 3 (Chapter 6) examined how syntactic complexity interacted with repetition accuracy across language groups (monolingual vs. bilingual) and clinical groups (TD vs. SLI), and whether variance in performance was mainly linked to language abilities (Klem et al., 2015), or to working memory capacities as estimated by simple and complex digit recall tasks (Delage et al., 2019, 2020; Zebib et al., 2020b).

To summarize, this dissertation aims to inform bilingual assessment practice by: 1) examining the clinical potential of heritage language assessment for identifying SLI in the German heritage-minority context, comparing standardized HL assessment to a LITMUS tool; 2) evaluating the clinical utility of the German LITMUS sentence- and nonword-repetition tasks and their robustness to language dominance; and 3) determining which SRT structure types best distinguish between typically developing children from those with SLI, and how this relates to computational complexity, general language ability and working memory across the different lingual and clinical groups. The extent to which these aims were achieved with the data collected and analyzed in this thesis is summarized in the subsections below.

7.2 Study 1: Evaluating the diagnostic value of heritage-language assessment for ruling in or out SLI in heritage bilinguals (Aim 1)

Building on Hamann et al. (2020) but shifting the focus to preschool- and primary-school-age heritage bilingual children acquiring Levantine Arabic in the German heritage context, Study 1 addressed the recommendation to assess heritage-language (HL) abilities alongside the L2 to avoid misdiagnosis of SLI in multilingual contexts. Study 1 extends Hamann et al. (2020) by comparing heritage BiTD children not only to late-successive BiTD children but also to heritage Arabic–German bilingual children with SLI and to monolingual, typically developing children acquiring Arabic in a majority context. As outlined in Chapter 4, heritage bilinguals show large individual variation in their HL skills due to reduced input conditions, and some interlanguage features overlap with markers of monolingual SLI. If the heritage language is not well developed in both children with and without SLI, its diagnostic value is limited. The study, therefore, investigated the feasibility of HL assessment and which diagnostic tools are suitable for heritage settings. Focusing on Levantine Arabic, spoken by a sizeable heritage minority in Germany, the study evaluated the clinical utility of one of the few available standardized tests for spoken Arabic, the ELO-L battery (Zebib et al., 2017), for ruling in or out SLI in simultaneous and early successive bilingual children acquiring Levantine Arabic as a heritage language in Germany. Given that LITMUS sentence repetition tasks have good diagnostic accuracy for distinguishing typical from impaired bilingual language development in the majority language, the study also asked whether a LITMUS sentence repetition task for spoken Arabic (Zebib et al., 2020a) developed for the Lebanese diglossic bilingual setting, in which Arabic enjoys a majority status, would be suitable for heritage-minority contexts and offer a faster, easy to administer alternative to standardized HL assessment. The study also examined how child-internal (age and WM abilities) and language environment factors affect typically developing heritage bilinguals' performance on the ELO-L subtests and the Arabic LITMUS SRT, to identify measures and language domains that are robust to reduced input conditions and could be exploited for diagnostic purposes. Finally, the study assessed whether all structures of the Arabic LITMUS-SRT are appropriate for evaluating HL morphosyntactic abilities, given the reported difficulties for heritage speakers with certain constructions, such as object relative clauses. To address these research questions, the study compared the performance

of heritage bilingual children with and without SLI to that of monolingual Arabic-speaking children and late successive bilingual children with typical language development as control groups on the two tests.

In line with Hamann et al. (2020), the results showed that the ELO-L test battery disadvantaged heritage bilingual children even when dominance-adjusted bilingual cutoff criteria were applied. Unlike monolingual and late successive bilingual children with typical development, who scored above the pathology cut-off, many heritage BiTD children (12/18) would be classified as BiSLI according to the norms in the test manual, especially on measures of expressive and receptive vocabulary and morphosyntactic production, which were negatively affected by early L2 exposure and current L2 use. Morphosyntactic comprehension and phonology, by contrast, appeared more resilient to reduced input: heritage bilinguals outperformed BiSLI children and did not differ significantly from their monolingual and late successive BiTD peers. This implies that relying solely on standardized L1 tests is not feasible in heritage contexts, and that HL assessment can help rule out but not confirm SLI in heritage children. Nevertheless, HL assessment can still be informative in the case of recently arrived late successive bilingual children with limited exposure to the societal language, such as children with forced displacement biographies. The latter children are expected to demonstrate stronger HL maintenance and not be disadvantaged by standardized assessment, since they spent an extended period in a monolingual, qualitatively rich language environment and likely had access to formal language registers.

The results for the LITMUS-SRT for Levantine Arabic, which was developed according to the LITMUS principles and includes structures with complexity aspects (embedding and *wh*-movement), known to be vulnerable in SLI but robust in bilingual language acquisition, were promising. In contrast to the ELO-L, heritage BiTDs showed comparable performance to their MoTD and late successive BiTD counterparts and significantly differed from the BiSLI group with large effect sizes, not only when the task was scored by “target structure”, but also for “identical repetition”. This is an encouraging result as this scoring measure is faster and easier than “target structure”. Moreover, the only language measure explaining variance in performance was the Morphosyntax Reception subtest of the ELO-L, on which heritage BiTD performed on par with their MoTD and late-successive BiTD peers. Importantly, the SRT task appeared to be robust against reduced input effects, as variance was not explained by any of the input variables,

indicating that the task is also fair for preschool and primary-school-age heritage bilingual children, replicating findings by Hamann et al. (2020) with older bilinguals, who also showed ceiling effects on the task. The results differed from those of Hamann et al., though, in that verbal short-term memory (as measured by FDS) significantly contributed to explaining the variance in the BiTD performance. This is probably due to the younger age of the bilingual participants in this sample, whose working memory capacities are still developing. This finding aligns with previous studies on LITMUS SRTs, confirming that processing computationally complex structures such as those in the Arabic LITMUS SRT requires greater working memory resources, which may be further taxed when bilingual children perform in their weaker language. All structures differentiated heritage BiTDs from BiSLIs, except object relative clauses. Although heritage BiTDs and BiSLIs did not differ significantly on this construction, their error patterns differed. BiSLIs produced mostly sentential fragments, suggesting difficulty with computational complexity (see also Study 3). BiTDs mainly omitted the resumptive object clitic or the relative pronoun within the relative clause, produced an alternative complex construction, or omitted the definite article of the relativized noun due to morphophonological properties of some words used in the stimuli. Overall, the results for the Arabic LITMUS SRT are very promising and show that the task can be administered as a first estimate of heritage-language abilities and can provide evidence of typical development in heritage and late successive bilingual children, which, in turn, would imply that their difficulties are unlikely to result from language impairment. At the same time, some structures, such as object relative clauses (though good indicators in monolingual contexts), should be avoided when designing tests for heritage bilinguals, and special attention should be paid to morphophonological features known to be vulnerable in heritage language acquisition.

It is important to acknowledge some limitations of the study. First, the BiSLI sample was relatively small ($n = 9$), precluding regression analyses. As a result, it was not possible to examine whether the effects of reduced HL input would be amplified in heritage BiSLI (Blom et al., 2019). The study would also have benefited from a control group of monolingual Arabic-speaking children with SLI. Including such a control group could shed light on the effects of SLI in monolingual versus bilingual contexts. Although the author attempted to recruit a MoSLI group in Palestine, the children recruited had crucial comorbidities, such as intellectual disability or unspecified disorders of psychological development (ICD-10: F89), so a reliable diagnosis with SLI was not

possible. Future research should replicate these findings with a larger, more diverse BiSLI sample in terms of severity and exposure patterns and include a MoSLI control group.

7.3 Study 2¹⁷⁶: Examining the diagnostic robustness of the German LITMUS sentence and nonword repetition tasks against the influence of language dominance (Aim 2)

Given the great variability in bilingual language exposure and the potential effect of language experience on language performance in bilingual children, Study 2 examined whether background variables related to bilingualism, particularly the degree of language dominance as measured by the relative amount of use and exposure, could compromise the diagnostic accuracy of the German LITMUS sentence repetition (Hamann et al., 2013) and quasi-universal nonword repetition tasks (Grimm, 2022; Grimm et al., 2014). As elaborated in 2.3.1 and 3.2.3, both tasks are part of the LITMUS toolbox and are proposed to successfully discriminate BiSLI from BiTD children by targeting language aspects of morphosyntactic and phonological complexity known to pose difficulties for monolingual children with SLI, but not for typically developing bilingual children (cf. Armon-Lotem et al., 2015a). Several studies investigating the diagnostic accuracy of the German LITMUS SRT report higher sensitivity and specificity when the task is scored using the more lenient "correct target structure" method, which assesses preservation of the target construction and disregards frequent bilingual errors such as lexical, case, and gender errors, provided they do not affect realization of the construction. For the QU-NWRT, better diagnostic accuracy has been reported for the phonologically more complex, language-dependent part of the task (Grimm, 2022). Hence, the study investigated whether combining the two tasks yields better diagnostic accuracy than single measures and whether a combination of sentence repetition scored by correct target structure and the structurally more complex language-dependent part of the QU-NWRT maximizes diagnostic accuracy. An innovative aspect of the study was the unbiased statistical method used to establish combined cut-off points separating TD from SLI in monolingual and bilingual children.

¹⁷⁶ The summary of Study 2 is taken from Abed Ibrahim and Fekete (2019).

To address the latter questions, an unsupervised machine learning algorithm, the Partitioning-Around-Medoids (PAM, Kaufman and Rousseeuw, 2009) was used for deriving a clinical category for the children as $\pm$ language-impaired based on their performance scores on SRT (scored by identical repetition and correct target structure) and QU-NWRT (global score and scores of the language independent and language dependent subtests) while withholding information about their clinical status based on standardized assessment in their HL and L2. Subsequently, diagnostic accuracy was calculated, and Firth penalized regression analyses were carried out to investigate which background variables (age of onset, length of exposure, degree of language dominance, socioeconomic status, and risk factors for SLI) best predicted clinical-group-membership yielded from the PAM-analysis based on the children's SRT and QU-NWRT performance scores. Results showed that although language dominance influenced the performance of bilingual typically developing children, especially on the SRT, the diagnostic accuracy of the tools was not compromised by dominance or socioeconomic status. Risk factors for SLI were significant predictors of clinical group membership in all models, while language dominance did not contribute to explaining membership in the TD versus SLI clusters for any combination of SRT and QU-NWRT measures. Moreover, the results confirmed that using a combination of tasks, each emphasizing a different aspect of language ability, yields better diagnostic accuracy than single measures. The combination was sensitive only to risk factors for SLI and not to language dominance or SES, a robustness not achieved by other tasks and procedures. As a final promising result, the study showed that using SRT_Tar together with the NWRT_LD produced the best diagnostic accuracy obtained so far for similarly constructed tasks. Thus, the study replicated the finding that the scoring method "correct target structure" is better suited than identical repetition for assessing the language abilities of bilingual children with German as L2. The significant contribution of the structurally more complex, language-dependent part of the QU-NWRT to diagnostic accuracy is consistent with robust effects of phonological complexity on SLI performance (e.g., Gallon et al., 2007; Ferré et al., 2012). The clinical implications clearly are that morphosyntactic and phonological complexity, especially in a combined measure, can serve as a reliable indicator of SLI in both monolingual and bilingual children.

7.4 Study 3: Examining the differential effects of syntactic complexity on SRT performance in monolingual and bilingual children with and without SLI, and their interaction with language and working-memory abilities (Aim 3)

As elaborated in 2.3.1.1, the construction of the German LITMUS-SRT was guided by the notion of syntactic complexity. Syntactically complex constructions are cross-linguistically vulnerable in children with SLI, who, given their impaired language faculty (associated with weaker, less stable linguistic representations in long-term memory and deficits in derivational operations) in addition to limited working memory capacity (Friedmann & Ruigendijk, 2017), experience particular difficulty with complex constructions. By contrast, complex syntax is a relative strength in typical bilingual language development and appears less susceptible to split-input effects than input-sensitive domains such as vocabulary and inflectional morphology, provided there is sufficient exposure to the L2 (Abed Ibrahim & Hamman, 2024; Hamann et al., 2017). Syntactically complex constructions thus represent strong candidates for distinguishing typical development from SLI and were incorporated into the German LITMUS sentence repetition task, with particular focus on clausal embedding, *wh*-movement, intervention (NP-restriction), and noncanonical word order (Friedmann et al., 2009; Hamann et al., 2007; Rizzi, 2004; Tuller et al., 2012).

The significant reliance of sentence repetition on language skills has been established in various studies, especially when stimulus length exceeds the phonological loop capacity (Klem et al., 2015; Moll et al., 2015; Riches et al., 2010). At the same time, increasing syntactic complexity is associated with greater processing load (Gibson, 1998; Gordon et al., 2001; Kimball, 1973), particularly for constructions with hierarchically non-local dependencies, such as *wh*-questions and relative clauses (Jakubowicz, 2011; van der Lely, 2005a, 2005b). Unlike children with SLI, children with typical language development are assumed to benefit from their well-developed linguistic representations, which help alleviate processing demands and facilitate the repetition of such constructions. Many comprehension and production studies, including sentence recall, report an interaction between WM and syntactic complexity (Arosio et al., 2012; Delage & Frauenfelder, 2019, 2020); some studies even show that different working-memory components (VSTM/VWM) differentially predict performance of monolingual and

bilingual children with and without SLI (e.g., Talli & Stavrakaki, 2020; Zebib et al., 2020b), implying that different mechanisms might be employed by children with SLI to perform the same task (Riches, 2020, p. 503). For this reason, in addition to verifying the diagnostic accuracy of the German LITMUS-SRT with a larger participant sample, Study 3 examined how monolingual and bilingual children with and without SLI differ in performance on constructions of varying syntactic complexity and how their performance is associated with WM capacities and linguistic ability. To this end, the 45 sentences of the German LITMUS-SRT were grouped by complexity (movement, embedding, and intervention as an extra factor), and regression analyses tested whether monolingual and bilingual children with and without SLI showed different associations between performance on complex constructions and linguistic and WM measures (forward digit span for VSTM; backward digit span for VWM).

The study's results replicated previous findings, showing the German LITMUS-SRT sharply distinguishes typical from impaired language development regardless of lingual status, with higher diagnostic accuracy for bilingual children when scoring is based on preservation of the target structure rather than identical repetition (Chilla et al., 2021; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018). Moreover, the cutoff associated with the highest diagnostic accuracy resembled that obtained in previous studies and closely approximated the cutoff derived via unsupervised machine learning in Study 2, with almost no overlap between typically developing bilinguals and monolingual or bilingual children with SLI, validating the robustness of the task.

In line with previous studies, repetition accuracy decreased in all groups as cumulative syntactic complexity increased, with much larger effect sizes in the MoSLI and BiSLI groups, who showed floor performance on structures containing two or more complexity factors (Level 2 & Level 3), indicating a total computational breakdown. These findings confirm the theoretical assumptions underlying the construction and classification of the sentences and show that aspects of computational complexity are particularly vulnerable in children with SLI, whose weak linguistic representations and limited working memory capacities are overtaxed by such structures.

Age-based analyses revealed significant differences between MoTDs and younger BiTD children for most Level 1 and Level 2 constructions, with the largest effect sizes for topicalizations and object relatives with intervention. Older BiTD children (primary school-age), on the other hand, showed a more homogeneous performance and patterned

like MoTDs. This finding aligns with Hamann et al. (2017), who reported that younger BiTD children still had difficulties with subject and object relatives (embedding and *wh*-movement), and long passives (smuggling). However, whether these difficulties reflect younger BiTD children's problems exploiting morphological case marking for theta-role assignment or stem from noncanonical word order cannot be determined from the present data. Nevertheless, although the younger BiTDs did not show mastery of the latter constructions, they still performed significantly better than the SLI groups, who performed very poorly on all of them. Thus, including such structures in a repetition task should allow reliable identification of SLI in both monolingual and bilingual children.

Comparable to the results in Ruigendijk and Friedmann (2017) and Hamann et al. (2017), all groups struggled substantially with Level 3, which comprised only object relatives with intervention; even MoTD children performed only slightly above chance. The greater difficulty with object relatives involving intervention cannot be attributed solely to noncanonical word order or case marking, since these features also occur in object relatives without intervention, which are repeated more accurately. An increased memory load through the presence of new referents (Gibson, 1998) is also unlikely to explain the greater difficulty repeating ORs with intervention, since these contain two lexical DPs, like most structures of Level 1 and Level 2, except for finite complement clauses (3 lexical DPs). This pattern supports Friedmann et al.'s (2009) analysis of the specific challenge posed by intervention in object relatives. The absence of observable intervention effects in the MoSLI and BiSLI groups likely reflects their general difficulty with all complex constructions (≥ 1 complexity factor), preventing the unique influence of intervention from manifesting.

Regarding the interaction of SRT performance with working memory capacities and linguistic ability, no association with working memory emerged for the SLI groups; their SRT performance was predicted by L2 morphosyntax, which seems at odds with previous findings (Delage & Frauenfelder, 2020; Zebib et al., 2020b). A possible explanation is that, in the SLI groups, most children scored around 3 on the forward digit span and near floor on the backward digit span, reducing interindividual variability and therefore the likelihood of detecting an association with performance. On the other hand, the results may reflect the special care taken in constructing the SRT stimuli to ensure complex constructions contained an equal number of lexical DPs. Where necessary, pronouns were inserted so that the number of lexical DPs per construction remained

comparable, ensuring that the number of lexical items that had to be held in the working-memory buffer was similar across complex constructions. This may also have prevented the influence of working memory from manifesting in children with SLI. The absence of working-memory effects in the SLI groups is a welcome result as it implies that the task taps “linguistic rather than domain-general working memory abilities” (Archibald & Harder Griebeling, 2016, p. 261f), consistent with its intended design. For the TD groups, MoTD performance on the task overall and on Level 1 structures was predicted only by vocabulary measures, whereas BiTD performance was predicted by forward digit span and L2 vocabulary. Interestingly, performance on the structurally more complex Level 2 was only predicted by FDS in both TD groups, confirming that processing of computationally more complex structures necessitates more WM capacities (Arosio et al., 2011; Jakubowicz, 2011). The absence of L2 morphosyntax effects can be ascribed to the lack of variance in the TD groups, since all participants reached the highest developmental level, namely, the ability to produce embedded constructions. BiTDs’ greater reliance on working memory for processing complex constructions may reflect weaker L2 vocabulary, a domain known to be vulnerable in bilinguals. While MoTDs can draw on more advanced language knowledge in long-term memory, BiTDs may compensate for weaker L2 skills by relying on intact working-memory capacities.

The study clearly showed that complex structures involving movement, embedding, and intervention reliably distinguish typical from impaired language development in both monolingual and bilingual children. Hence, tasks incorporating such complex constructions offer promising tools for establishing typical language development in bilingual children. However, the study also showed that the most complex constructions, such as object relatives with intervention and topicalizations, do not reliably distinguish SLI from typical development, since many BiTD children in the studied age group also showed floor performance. Nevertheless, performance on the SRT as a whole, which also includes simpler constructions, achieved the desired separation with no overlap between BiTD and MoSLI or BiSLI. Thus, using a range of structures that vary in complexity appears promising for clinical purposes, as it is less frustrating for children with low language ability. These results are important for developing shorter versions of the German LITMUS SRT, which would be welcomed by clinicians and educators, given the limited time available for language assessment.

7.5 Conclusion and outlook

The studies presented in this thesis provide a thorough evaluation of the clinical potential of standardized L1 assessment and of assessment using LITMUS sentence and nonword repetition tasks for identifying SLI in preschool and primary-school-age (ages: 5;6-8;11 years) monolingual, 2L1, and early L2 bilingual children acquiring German as a societal language. In light of the findings of the three studies, it can be concluded that standardized L1 assessment, like standardized L2 assessment, poses a high risk of overidentifying heritage BiTD children as language-impaired, even when pathology cut-offs are adjusted for language dominance. By contrast, owing to their careful construction and cross-linguistic availability, LITMUS repetition tools provide time-efficient, easy-to-administer measures that reliably distinguish typical language development from SLI in heritage bilingual children, not only in the societal language (L2) but also in the heritage/minority language (HL/L1).

In Study 1, results showed that reduced HL input and possibly attrition can cause diagnostic confounds when heritage bilingual children are tested with HL standardized assessment tools. This is particularly true for input-sensitive domains like vocabulary and expressive morphosyntax, whereas other domains, such as HL phonology and receptive vocabulary, appear relatively robust to reduced input and can therefore be exploited for diagnostic purposes. As for HL vocabulary, it should be noted that the study examined only vocabulary breadth, which is vulnerable in both BiTD and SLI. Because breadth appears particularly affected in heritage contexts, it may be clinically more informative to assess vocabulary depth, that is, quality rather than quantity (see Armon-Lotem et al., 2021). With respect to complex morphosyntax, the study showed that the LITMUS sentence-repetition task for Levantine Arabic, grounded in the notion of computational complexity, offers a promising tool for distinguishing BiTD from BiSLI when administered in the heritage language, corroborating previous findings on the robustness of complex syntax in bilingual typical development and its vulnerability in SLI. For late-successive bilingual children with limited L2 exposure, the study showed that HL assessment can be particularly informative, since these children are likely to exhibit strong HL maintenance after extended exposure to a monolingual, qualitatively rich language environment with access to formal registers. By contrast, heritage BiSLI children appear to be at substantial risk of losing their HL (Blom et al., 2019, 2022), not

because of cumulative effects of bilingualism and SLI but because they lack opportunities to reach their full potential in the HL “following their own trajectory of language development” (Novogrodsky & Meir, 2022, p. 22). This underscores the importance of bilingual awareness programs that inform parents about the need to continue providing HL input to their bilingual children. Overall, the results show that assessing HL abilities alongside the societal language can inform clinical decisions; however, caution is required when selecting HL assessment tools.

Results on the diagnostic accuracy of the German LITMUS SRT and QU-NWRT reported in Study 2 are particularly promising, not least because they emerged from unsupervised PAM clustering. A key advantage of PAM clustering is that classification as TD or SLI is based solely on children’s performance scores on the repetition tasks while their clinical status is withheld, thereby mitigating pre-selection bias and avoiding the classical circularity between classification and verification (Redmond et al., 2019). A further advantage of the PAM algorithm is that it allowed the examination of the combined diagnostic power of sentence repetition and nonword repetition, thereby enabling the use of combined clinical-marker tasks (e.g., Bonifacci et al., 2020; Conti-Ramsden, 2003). The results are encouraging for German and the procedures applied in this study: the task reliably distinguishes SLI from typical development in monolingual and bilingual children, irrespective of socioeconomic status or degree of language dominance. Although the tasks showed good diagnostic accuracy when considered separately, some bilingual children were still overidentified by single SRT and NWRT measures. By contrast, combining the two tasks, each targeting a different aspect of language ability, reduced misdiagnoses and improved diagnostic accuracy. Notably, the combination of the SRT “correct target structure” score and the structurally more complex, language-dependent part of the QU-NWRT yielded the highest accuracy and was predicted only by risk factors for SLI as estimated by the parental questionnaire PaBiQ, confirming that children with SLI have exceptional difficulties with morphosyntactic and phonological complexity. This highlights the superiority of combined measures over single measures for reliable diagnosis, as combined measures are more likely to account for the heterogeneity of bilingualism and SLI. Hence, a crucial clinical implication is not to use the tasks in isolation and complement them with the parental questionnaire for bilingual children.

Building on previous BiLaD project research (e.g., Abed Ibrahim et al., 2018; Hamann et al., 2017; Zebib et al., 2020b) and Study 2, Study 3 examined the potential of complex syntactic constructions to identify SLI in bilingual populations and their interaction with language and working-memory abilities. The study replicated the high diagnostic accuracy of the German LITMUS-SRT for monolingual and bilingual children, showing that an SRT incorporating movement, embedding, and intervention as complexity factors sharply distinguishes typical from impaired language development regardless of children's lingual status. As in other studies using LITMUS SRTs (e.g., Armon-Lotem & Meir, 2016; Fleckstein et al., 2018), the cut-off scores separating TD from SLI are lower for bilinguals than for monolinguals, indicating that monolingual norms do not apply even to carefully constructed tests such as the LITMUS SRT. Importantly, at the group level, although MoTDs outperformed BiTDs, there was no overlap between BiTDs and either SLI group. BiTDs performed significantly better than both MoSLIs and BiSLIs, while BiSLIs and MoSLIs showed comparable profiles. The similar performance of BiSLIs and MoSLIs suggests that, contrary to the cumulative-effects hypothesis, no additive effect of bilingualism and SLI was evident for complex morphosyntax, at least in the age range and language combinations investigated in this thesis. Importantly, repetition accuracy was associated with German standardized test scores, confirming the task's concurrent validity.

The data showed that children with SLI differed from TD children on all constructions, with the magnitude of differences amplified for computationally more complex constructions. Although this is beyond the scope of the dissertation and the thesis does not attempt to adopt a particular theoretical framework, it is interesting to speculate about the source of the disproportionate difficulties children with SLI encounter across structures of different complexity levels. In this study, computational complexity is defined in generative terms, based on the number and nature of syntactic operations involved in deriving a structure (Jakubowicz, 2005, 2011; Tuller et al., 2012; van der Lely, 2005a, 2005b), rather than in construction grammar or usage-based terms, which define complexity by length, frequency, structural similarity, and the presence of nested structures (Diessel, 2009; Kidd et al., 2007; Lieven, 2011; Tomasello, 2003). The former appears to better characterize the sentences of the German LITMUS SRT, which include nested structures that control for animacy, number, and length effects but differ in computational complexity, such as subject versus object relative clauses. Usage-based

approaches argue that the use and retention of complex structures depend on discourse regularities and distributional frequency in the input, and some studies within this framework attribute SLI children's difficulties to the lower frequency or avoidance of complex constructions in the input directed at them compared to TD children as a result of context optimization, that is, caregivers adjust input to a child's abilities by using simpler constructions (e.g., Prigent et al., 2015). Although this could predict that children with SLI avoid or use fewer complex constructions in general, much structure learning through statistical probabilities also depends on the ability to process sequences, so input frequency alone does not fully explain the differential difficulties children with SLI encounter across SRT complexity levels. Additionally, the finding that repetition accuracy is not, or not solely, predicted by working memory measures in the TD and SLI groups indicates that repetition does not merely rely on working memory resources; this implies underlying deficits in the grammatical mechanisms responsible for computing non-local dependencies and thus provides some support for domain-specific accounts of SLI. Nevertheless, the data presented in Study 3 do not preclude interaction effects between structural complexity and input frequency across the linguistic and clinical groups (see Adani et al., 2017). To address this question directly, further research that carefully controls for these variables and employs advanced statistical procedures is necessary. Additionally, future studies should examine the tasks' applicability to other clinical populations, for example, orally trained hard-of-hearing children (cf. Abed Ibrahim et al., 2025; Chilla & Abed Ibrahim, 2026).

Despite the promising results, it is also important to acknowledge a number of limitations to the studies in this thesis. A crucial limitation is the small sample size of children with SLI, partly resulting from the strict clinical verification procedure used in the BiLaD project. This procedure required two language domains per language to fall below pathology cutoffs for classification as SLI and led to the reclassification of a significant proportion of children in language therapy as typically developing, suggesting prior overdiagnosis and possibly biasing the retained BiSLI sample toward more severe cases. Another limitation is that the SLI group comprised mainly simultaneous bilinguals who were relatively dominant in German, which constrained the investigation of bilingualism and SLI effects on heritage language development. Future research should replicate these findings with a larger, more heterogeneous sample of BiSLI children varying in severity and language-exposure patterns, and should include not only 2L1 and

early L2 learners but also late successive L2 bilinguals. It would be valuable to examine the stability of diagnostic accuracy longitudinally and to examine HL and L2 cross-language interaction effects within the same bilingual participants, which is feasible given the available HL and L2 data.

Despite the aforementioned limitations (and others I may have missed), this thesis provides novel data and useful insights for developing diagnostic tools that account for the strengths and weaknesses of bilingual children acquiring German as the societal language. The German LITMUS sentence and nonword repetition tasks, especially when combined, showed good accuracy in distinguishing preschool- and primary-school-age 2L1 and early L2 bilingual children with SLI from same-age typically developing peers. Clinicians may therefore consider these tasks for inclusion in screening protocols; however, Studies 2 and 3 indicate further steps are needed before clinical implementation. For easier administration and greater time efficiency, the first step is to shorten the tasks, since the pilot versions are too long and time-consuming for routine screening in educational and clinical settings. For the SRT, Study 3 indicates that it is necessary to eliminate topicalizations and object relatives with intervention and to carefully evaluate the role of case-marked DPs for thematic role assignment, as these may be sources of difficulty, especially for younger bilingual children. In Study 2, age predicted NWRT global scores in addition to SLI risk factors, and in Study 3, school-age children outperformed preschoolers on complex LITMUS SRT constructions. These results suggest diagnostic accuracy may vary with age and length of exposure to the L2 (LoE). Therefore, establishing age- and LoE-specific cutoffs may be necessary to optimize classification accuracy. As to the LITMUS QU-NWRT, a shortened version was developed toward the end of this dissertation by the test authors (see Grimm, 2022) and is currently being piloted with a large cohort of bilingual children (see Gagarina et al., 2026). The German LITMUS SRT is also being shortened by the authors and is currently being evaluated on a larger sample of monolingual and bilingual children within the EVA-SRT project (Evaluation von LITMUS-Satzwiederholungsverfahren für die sprachdiagnostische Praxis; DOI 10.17605/OSF.IO/8KB4N).

With respect to the newly developed LITMUS repetition tools, this dissertation provided evidence for the clinical validity of the LITMUS QU-NWRT and SRT, which target aspects of phonological and syntactic complexity that are vulnerable in SLI but robust in typical bilingual acquisition. From a theoretical perspective, the findings

indicate that SLI children's deficits are not limited to working memory. Despite keeping memory load constant in terms of length across different stimuli, MoSLI and BiLSI children struggled disproportionately with structurally more complex stimuli of equal length, suggesting the deficit is linguistic in nature and rooted in weaker, less stable linguistic representations that working memory alone cannot compensate for. Moreover, taken together, the studies in this dissertation showed that for phonology and complex syntax, despite quantitative and qualitative interindividual variation, bilingualism does not overburden children, including those with SLI, who are capable of acquiring two languages "following their own trajectory of language development" (Novogrodsky & Meir, 2022, p. 22). At the same time, BiSLI children risk losing their heritage language if input in that language is discontinued following uninformed recommendations to only use the societal language. From an applied perspective, the findings show that, whether one adopts a domain-specific or domain-general explanation, difficulties with phonological complexity and with computationally complex constructions involving (hierarchical) non-local dependencies can serve as reliable clinical markers for flagging monolingual and bilingual children at risk of SLI and in need of differential diagnosis, particularly when used in combination. As such, LITMUS repetition tasks featuring these constructions offer reliable screening tools for SLI that can help protect against pathologizing bilingualism by directing the limited clinical resources toward children who truly need language intervention services.

References

- Abdalla, F. (2002). *Specific language impairment in Arabic-speaking children: deficits in morphosyntax* [Doctoral dissertation, McGill University]. <https://escholarship.mcgill.ca/concern/theses/z029p522x>
- Abdalla, F. (2011). *Language Impairment in Arabic: Deficits in Morphosyntax*. VDM Verlag Dr. Müller
- Abdalla, F., Aljenaie, K., & Mahfoudhi, A. (2013). Plural noun inflection in Kuwaiti Arabic-speaking children with and without Specific Language Impairment. *Journal of Child Language*, 40(1), 139-168. <https://doi.org/10.1017/S0305000912000499>
- Abdalla, F., & Crago, M. (2008). Verb morphology deficits in Arabic-speaking children with specific language impairment. *Applied Psycholinguistics*, 29(2), 315-340. <https://doi.org/10.1017/S0142716408080156>
- Abed Ibrahim, L. (2025). Assessment of Heritage Language Abilities in Bilingual Arabic-German Children with and without Developmental Language Disorder: Comparing a Standardized Test Battery for Spoken Arabic to an Arabic LITMUS Sentence Repetition Task. *Journal of Speech Language and Hearing Research*, 68(3S), 1569-1601. https://doi.org/10.1044/2024_jslhr-23-00617
- Abed Ibrahim, L., Chilla, S., Hänel-Faulhaber, B. (2025). Developmental Trajectories of German as a Spoken Language in Typical and Hard of Hearing Children with Forced Displacement Background: A Longitudinal Pilot Study. In A. Yedetore, R. Bonney & Y. Zhang, (Eds.), *Proceedings of BUCLD 49* (pp. 19-34). Somerville: Cascadilla Press.
- Abed Ibrahim, L., & Fekete, I. (2019). What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks. *Frontiers in Psychology*, 9, 2757. <https://doi.org/10.3389/fpsyg.2018.02757>
- Abed Ibrahim, L., & Hamann, C. (2017). Bilingual Arabic-German & Turkish-German children with and without specific language impairment: comparing performance in sentence and Nonword repetition tasks. In Maria LaMendola & Jennifer Scott (Eds.), *Proceedings of BUCLD 41* (pp. 1-17). Somerville, MA: Cascadilla Press.
- Abed Ibrahim, L., & Hamann, C. (2024). German Litmus Sentence-Repetition Task as a screening tool for the identification of SLI in bilingual children in Germany: The role of syntactic complexity and working memory. In H. A. A. AlThagafi & J. Ray (Eds.), *Proceedings of BUCLD 48* (pp. 1-16). Cascadilla Press.
- Abed Ibrahim, L., Hamann, C., & Fekete, I. (2020). Language assessment of bilingual Arabic-German heritage and refugee children: Comparing performance on LITMUS repetition tasks. In M. M. Brown & A. Kohut (Eds.), *Proceedings of BUCLD 44* (pp. 1-17). Somerville: Cascadilla Press.
- Abed Ibrahim, L., Hamann, C., & Öwerdieck, D. (2018). Identifying specific language impairment (SLI) across different bilingual populations: German sentence repetition task (SRT). In A. B. Bertolini & M. J. Kaplan (Eds.), *Proceedings of BUCLD 42* (pp. 1-14). Somerville, MA: Cascadilla Press.
- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59(2), 249-306. <https://doi.org/10.1111/j.1467-9922.2009.00507.x>
- Abutalebi, J., & Rietbergen, M. J. (2014). Neuroplasticity of the bilingual brain: Cognitive control and reserve. *Applied Psycholinguistics*, 35(5), 895-899. <https://doi.org/10.1017/S0142716414000186>
- Abutbul-Oz, H., Armon-Lotem, S., & Walters, J. (2012). *Bilingual Parents Questionnaire (BIPAQ)*. Unpublished instrument. English Department, Bar Ilan University, Ramat Gan, Israel.
- Adani, F. (2011). Rethinking the acquisition of relative clauses in Italian: towards a grammatically based account. *Journal of Child Language*, 38(1), 141-165. <https://doi.org/10.1017/S0305000909990250>
- Adani, F. (2012). Some Notes on the Acquisition of Relative Clauses: New Data and Open Questions. *Enjoy Linguistics! Papers offered to Luigi Rizzi on the occasion of his 60th birthday*, 6-13. Internet celebration, CISCL Siena.
- Adani, F., Forgiarini, M., Guasti, M. T., & van der Lely, H. K. (2014). Number dissimilarities facilitate the

- comprehension of relative clauses in children with (Grammatical) Specific Language Impairment. *Journal of Child Language*, 41(4), 811-841. <https://doi.org/10.1017/S0305000913000184>
- Adani, F., Stegenwallner-Schütz, M., Haendler, Y., & Zukowski, A. (2016). Elicited production of relative clauses in German: Evidence from typically developing children and children with specific language impairment. *First Language*, 36(3), 203-227. <https://doi.org/10.1177/0142723716648842>
- Adani, F., Stegenwallner-Schütz, M., & Niesel, T. (2017). The peaceful co-existence of input frequency and structural intervention effects on the comprehension of complex sentences in German-speaking children. *Frontiers in Psychology*, 8, 1590. <https://doi.org/10.3389/fpsyg.2017.01590>
- Adani, F., van der Lely, H. K., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua. International review of general linguistics. Revue internationale de linguistique generale*, 120(9-3), 2148-2166. <https://doi.org/10.1016/j.lingua.2010.03.018>
- Adi-Japha, E., Berberich-Artzi, J., & Libnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child development*, 81(5), 1356-1366. <https://doi.org/10.1111/j.1467-8624.2010.01477.x>
- Aguilar-Mediavilla, E., Sanz-Torrent, M., & Serra-Raventós, M. (2007). Influence of phonology on morpho-syntax in Romance languages in children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, 42(3), 325-347. <https://doi.org/10.1080/13682820600881527>
- Ahmad Rusli, Y., & Montgomery, J. W. (2017). Children's Comprehension of Object Relative Sentences: It's Extant Language Knowledge That Matters, Not Domain-General Working Memory. *Journal of Speech, Language, and Hearing Research*, 60(10), 2865-2878. https://doi.org/10.1044/2017_JSLHR-L-16-0422
- Akıncı, M. A., Pfaff, C.W., & Dollnick, M. (2013). Orthographic and morphological aspects of written Turkish in France, Germany and Turkey. In I. Ergenc (Ed.), *Proceedings of the Turkish linguistic conference, Antalya 2008* (pp. 363-372). Wiesbaden: Otto Harrassowitz Verlag.
- Albirini, A. (2014). Toward understanding the variability in the language proficiencies of Arabic heritage speakers. *International Journal of Bilingualism*, 18(6), 730-765. <https://doi.org/10.1177/1367006912472>
- Albirini A. (2015). Factors affecting the acquisition of plural morphology in Jordanian Arabic. *Journal of Child Language*, 42(4), 734-762. <https://doi.org/10.1017/S0305000914000270>
- Albirini, A. (2016). *Modern Arabic Sociolinguistics. Diglossia, variation, codeswitching, attitudes and identity*. New York: Routledge. <https://doi.org/10.4324/9781315683737>
- Albirini, A. (2018). The Role of Age of Exposure to English in the Development of Arabic as a Heritage Language in the United States. *Language Acquisition*, 25(2), 178-196. <https://doi.org/10.1080/10489223.2016.1234620>
- Albirini, A., & Benmamoun, E. (2014). Aspects of second-language transfer in the oral production of Egyptian and Palestinian heritage speakers. *International Journal of Bilingualism*, 18(3), 244-273. <https://doi.org/10.1177/1367006912441>
- Albirini, A., & Benmamoun, E. (2022). Arabic Diglossia and Heritage Arabic Speakers. In E. Saiegh-Haddad, L. Laks & C. McBride (Eds.), *Handbook of Literacy in Diglossia and in Dialectal Contexts. Literacy Studies*, (vol. 22, pp. 361-379). Springer. https://doi.org/10.1007/978-3-030-80072-7_15
- Albirini, A., Benmamoun, E., & Chakrani, B. (2013). Gender and number agreement in the oral production of Arabic Heritage speakers. *Bilingualism: Language and Cognition*, 16(1), 1-18. <https://doi.org/10.1017/S1366728912000132>
- Albirini, A., Benmamoun, E., & Saadah, E. (2011). Grammatical Features of Egyptian and Palestinian Arabic Heritage Speakers' Oral Production. *Studies in Second Language Acquisition*, 33(2), 273-303. <https://doi.org/10.1017/S0272263110000768>
- Albrecht, K. M. (2017). *The Identification of Typical and Atypical Phonological Acquisition in Turkish-German Bilingual Children* [Doctoral Dissertation, University of Sheffield].
- Aljenaie, K. (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child*

- Language*, 37(4), 841-863. <https://doi.org/10.1017/S0305000909990031>
- Alkhudidi, A., Hakooz, Y., Walker, M., Stevenson, R., & Rafat, Y. (2020). 8. The production of marked Arabic consonants by Arabic-English bilingual children living in Canada. In E. Babatsouli & M. Ball (Eds.), *An anthology of bilingual child phonology* (pp. 187-208). Multilingual Matters. <https://doi.org/10.2307/jj.22730686.12>
- Allen, R. J., & Baddeley, A. D. (2009). Working memory and sentence recall. In A. S. C. Thorn & M. P. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 63-85). Psychology Press. <https://doi.org/10.4324/9780203938966>
- Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87(2), 85-106. <https://doi.org/10.1016/j.jecp.2003.10.002>
- Alloway, T. P., & Gathercole, S. E. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology*, 17(2), 207-220. <https://doi.org/10.1080/095414404400000005>
- Al-Momani, I. (2015). Overt and null subject pronouns in Jordanian Arabic. *Advances in Language and Literary Studies*, 6(4), 1-10. <http://dx.doi.org/10.7575/aiaa.all.s.v.6n.4p.1>
- Al Omary, M. (2024). *Arabic Heritage Speakers' Production and Perception of Arabic Emphatic-Plain Contrasts* [Doctoral dissertation, The University of Wisconsin-Milwaukee]. UW Milwaukee Electronic Theses and Dissertations. <http://digital.library.wisc.edu/1793/93500>
- Alsudais, A., Alotaibi, W., & Alomary, F. (2022). Similarities between Arabic dialects: Investigating geographical proximity. *Information Processing & Management*, 59(1), Article 102770. <https://doi.org/10.1016/j.ipm.2021.102770>
- Alt M. (2011). Phonological working memory impairments in children with specific language impairment: where does the problem lie?. *Journal of Communication Disorders*, 44(2), 173-185. <https://doi.org/10.1016/j.jcomdis.2010.09.003>
- Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 49(5), 941-954. [https://doi.org/10.1044/1092-4388\(2006/068\)](https://doi.org/10.1044/1092-4388(2006/068))
- Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research*, 47(2), 407-420. [https://doi.org/10.1044/1092-4388\(2004/033\)](https://doi.org/10.1044/1092-4388(2004/033))
- Alt, M., & Spaulding, T. (2011). The effect of time on word learning: an examination of decay of the memory trace and vocal rehearsal in children with and without specific language impairment. *Journal of Communication Disorders*, 44(6), 640-654. <https://doi.org/10.1016/j.jcomdis.2011.07.001>
- Altman, C., Harel, E., Meir, N., Iluz-Cohen, P., Walters, J., & Armon-Lotem, S. (2022). Using a monolingual screening test for assessing bilingual children. *Clinical Linguistics & Phonetics*, 36(12), 1132-1152. <https://doi.org/10.1080/02699206.2021.2000644>
- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(2), 239-273. <https://doi.org/10.1017/S030500091400049X>
- Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511975073>
- Ambridge, B., Rowland, C. F., & Pine, J. M. (2008). Is structure dependence an innate constraint? New experimental evidence from children's complex-question production. *Cognitive Science*, 32(1), 222-255. <https://doi.org/10.1080/03640210701703766>
- American Speech-Language-Hearing Association. (2023). *Ad Hoc Committee on Bilingual Service Delivery: Competencies, expectations, and recommendations for multilingual service delivery [Final report]*. <https://www.asha.org/siteassets/reports/ahc-bilingual-service-delivery-final-report.pdf> [PDF]
- Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring and classification accuracy of

- vocabulary testing in bilingual children. *Language, Speech, and Hearing Services in Schools*, 49(1), 85-97. https://doi.org/10.1044/2017_LSHSS-16-0081
- Andersson, K., Hansson, K., Rosqvist, I., Lyberg Åhlander, V., Sahlén, B., & Sandgren, O. (2019). The contribution of bilingualism, parental education, and school characteristics to performance on the Clinical Evaluation of Language Fundamentals: Swedish. *Frontiers in psychology*, 10, 1586. <https://doi.org/10.3389/fpsyg.2019.01586>
- Andreou, M., Dosi, I., Papadopoulou, D., & Tsimpli, I. M. (2020). Heritage and non-heritage bilinguals: The role of biliteracy and bilingual education. In B. Brehmer & J. Treffers-Daller (Eds.), *Lost in transmission: The role of attrition in heritage language development* (pp. 172-196). Amsterdam: John Benjamins. <https://doi.org/10.1075/sibil.59.07and?locatt=mode:legacy>
- Andres, E. M., Earnest, K. K., Zhong, C., Rice, M. L., & Raza, M. H. (2021). Family-Based Whole-Exome Analysis of Specific Language Impairment (SLI) Identifies Rare Variants in BUD13, a Component of the Retention and Splicing (RES) Complex. *Brain Sciences*, 12(1), 47. <https://doi.org/10.3390/brainsci12010047>
- Andrés-Roqueta, C., & Katsos, N. (2020). A distinction between linguistic and social pragmatics helps the precise characterization of pragmatic challenges in children with autism spectrum disorders and developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(5), 1494-1508. https://doi.org/10.1044/2020_JSLHR-19-00263
- Anthony, J.L., Williams, J.M., & Aghara, R.G. et al. (2010). Assessment of individual differences in phonological representation. *Reading and Writing: An interdisciplinary Journal*, 23(8), 969-994. <https://doi.org/10.1007/s11145-009-9185-7>
- Aoun, J. (2000). Resumption and last resort. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada [Documentation of Studies in Theoretical and Applied Linguistics]*, 16(spe), 13-43. <https://doi.org/10.1590/S0102-44502000000300001>
- Aoun, J., Benmamoun, E., & Choueiri, L. (2010). *The Syntax of Arabic*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511691775>
- Aoun, J., & Choueiri, L. (1996). *Resumptive epithets*. Ms., University of Southern California, Los Angeles, California.
- Aoun, J., Choueiri, L., & Hornstein, N. (2001). Resumption, Movement, and Derivational Economy. *Linguistic Inquiry*, 32(3), 371-403. <https://doi.org/10.1162/002438901750372504>
- Archibald, L. M. (2017). Working memory and language learning: A review. *Child Language Teaching and Therapy*, 33(1), 5-17. <https://doi.org/10.1177/026565901665420>
- Archibald, L. M. (2018). The Reciprocal Influences of Working Memory and Linguistic Knowledge on Language Performance: Considerations for the Assessment of Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, 49(3), 424-433. https://doi.org/10.1044/2018_LSHSS-17-0094
- Archibald, L. M., & Gathercole, S. E. (2006). Nonword repetition: a comparison of tests. *Journal of Speech, Language, and Hearing Research*, 49(5), 970-983. [https://doi.org/10.1044/1092-4388\(2006/070\)](https://doi.org/10.1044/1092-4388(2006/070))
- Archibald, L. M., & Gathercole, S. E. (2007). The complexities of complex memory span: Storage and processing deficits in specific language impairment. *Journal of Memory and Language*, 57(2), 177-194. <https://doi.org/10.1016/j.jml.2006.11.004>
- Archibald, L. M., & Harder Griebeling, K. (2016). Rethinking the connection between working memory and language impairment. *International Journal of Language & Communication Disorders*, 51(3), 252-264. <https://doi.org/10.1111/1460-6984.12202>
- Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52(4), 899-914. [https://doi.org/10.1044/1092-4388\(2009/08-0099\)](https://doi.org/10.1044/1092-4388(2009/08-0099))
- Archibald, L. M., Joanisse, M. F., & Munson, B. (2013). Motor Control and Nonword Repetition in Specific Working Memory Impairment and SLI. *Topics in language disorders*, 33(3), 255-267. <https://doi.org/10.1097/TLD.0b013e31829cf5e7>

- Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, 10(1), 79-99. <https://doi.org/10.1017/S1366728906002835>
- Arias, G., & Friberg, J. (2017). Bilingual Language Assessment: Contemporary Versus Recommended Practice in American Schools. *Language, Speech, and Hearing Services in Schools*, 48(1), 1-15. https://doi.org/10.1044/2016_LSHSS-15-0090
- Armon-Lotem, S. (1996). What early Hebrew words teach us about root infinitives. In C. Koster & F. Wijnen (Eds.), *Proceedings of 1995 GALA Conference* (pp. 56-67), University of Groningen.
- Armon-Lotem, S. (2005). The Acquisition of Subordination: From Preconjunctionals to Later Use. In: D. D. Ravid & H.BZ. Shyldkrot (Eds.), *Perspectives on Language and Language Development* (pp. 191-202). Boston, MA: Springer. https://doi.org/10.1007/1-4020-7911-7_15
- Armon-Lotem, S. (2010). Instructive bilingualism: Can bilingual children with specific language impairment rely on one language in learning a second one? *Applied Psycholinguistics*, 31(2), 253-260. <https://doi.org/10.1017/S0142716409990385>
- Armon-Lotem S. (2014). Between L2 and SLI: inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. *Journal of Child Language*, 41(1), 3-33. <https://doi.org/10.1017/S0305000912000487>
- Armon-Lotem, S. (2018). SLI in bilingual development. In A. V. Bar-On & D. Ravid (Eds.), *Handbook of communication disorders* (pp. 617-641). De Gruyter. <https://doi.org/10.1515/9781614514909-031>
- Armon-Lotem, S., Danon, G., & Walters, J. (2008). The use of prepositions by bilingual SLI children: The relative contribution of representation and processing. In A. Gavarró & M. João Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007* (pp. 41-46). Cambridge: Cambridge Scholars Press. <https://doi.org/10.65325/EB885>
- Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.) (2015a). *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters.
- Armon-Lotem, S., & Grohmann, K. K. (2021). *Language impairment in multilingual settings: LITMUS in action across Europe*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tilar.29>
- Armon-Lotem, S., Haman, E., Jensen de López, K., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., et al. (2015b). A large-scale cross-linguistic investigation of the acquisition of passive. *Language Acquisition*, 23(1), 27-56. <https://doi.org/10.1080/10489223.2015.1047095>
- Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51(6), 715-731. <https://doi.org/10.1111/1460-6984.12242>
- Armon-Lotem, S., & Meir, N. (2018). The Nature of Exposure and Input in Early Bilingualism. In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 193-212). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316831922.011>
- Armon-Lotem, S., Rose, K., & Altman, C. (2021). The development of English as a heritage language: The role of chronological age and age of onset of bilingualism. *First Language*, 41(1), 67-89. <https://doi.org/10.1177/01427237209298>
- Armon-Lotem, S., Walters, J., & Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguistic Approaches to Bilingualism*, 1(3), 291-317. <https://doi.org/10.1075/lab.1.3.04arm>
- Arosio, F., Guasti, M. T., & Stucchi, N. (2011). Disambiguating information and memory resources in children's processing of Italian relative clauses. *Journal of psycholinguistic research*, 40, 137-154. <https://doi.org/10.1007/s10936-010-9160-0>
- Arosio, F., Yatsushiro, K., Forgiarini, M., & Guasti, M. T. (2012). Morphological Information and Memory Resources in Children's Processing of Relative Clauses in German. *Language Learning and Development*, 8(4), 340-364. <https://doi.org/10.1080/15475441.2011.634691>

- Audring, J. (2014). Gender as a complex feature. *Language Sciences*, 43, 5-17. <https://doi.org/10.1016/j.langsci.2013.10.003>
- Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, 68(6), 627-636. <https://doi.org/10.1016/j.jclinepi.2014.12.014>
- Autorengruppe Bildungsberichterstattung (2010). *Bildung in Deutschland 2010. Ein indikatorengestützter Bericht mit einer Analyse zu Perspektiven des Bildungswesens im demografischen Wandel*. Berlin: Bundesministerium für Bildung und Forschung.
- Authoring Group Educational Reporting (2016). *Education in Germany 2016. An indicator-based report including an analysis of education and migration*. Bielefeld: wbv. <https://www.bildungsbericht.de/en/au-thoring-group>
- Avram, L., Coene, M., & Sevcenco, A. (2015). Theoretical implications of children's early production of Romanian Accusative clitics. *Lingua*, 161, 48-66. <https://doi.org/10.1016/j.lingua.2014.03.004>
- Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's Aphasics. In Y. Grodzinsky, L. P. Shapiro & D. Swinney (Eds.), *Language and the Brain* (pp. 295-313). New York: Academic Press. <https://doi.org/10.1016/B978-012304260-6/50017-7>
- Avrutin, S., & Wexler, K. (1992). Development of Principle B in Russian: Coindexation at LF and coreference. *Language Acquisition*, 2, 259-306. https://doi.org/10.1207/s15327817la0204_2
- Babatsouli, E. (2021). Diversity considerations in speech and language disorders: a focus on training. In J. S. Damico, N. Müller and M. J. Ball (Eds.), *The Handbook of Language and Speech Disorders*, 2nd ed. (pp. 33-52). MA: Wiley-Blackwell. <https://doi.org/10.1002/9781119606987.ch2>
- Babatsouli, E. (2024). Phonological Disorders in Child Bilingualism. In M. Amengual (Ed.), *The Cambridge Handbook of Bilingual Phonetics and Phonology* (pp. 308-338). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009105767.015>
- Badawi, E. S., Carter, M. G., & Gully, A. (2004). *Modern Written Arabic. A Comprehensive Grammar*. Londres: Routledge.
- Baddeley A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, 4(11), 417-423. [https://doi.org/10.1016/s1364-6613\(00\)01538-2](https://doi.org/10.1016/s1364-6613(00)01538-2)
- Baddeley A. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36(3), 189-208. [https://doi.org/10.1016/s0021-9924\(03\)00019-4](https://doi.org/10.1016/s0021-9924(03)00019-4)
- Baddeley A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, 63, 1-29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, 105(1), 158-173. <https://doi.org/10.1037/0033-295x.105.1.158>
- Baddeley, A., & Hitch, G. J. (1974). Working Memory. *Psychology of Learning and Motivation*, 8, 47-89. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baddeley, A., Hitch, G. J., & Allen, R. (2020). A multicomponent model of working memory. In R. Logie, V. Camos & N. Cowan (Eds.), *Working memory: State of the science* (pp. 10-43). Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198842286.003.0002>
- Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, 61(3), 438-456. <https://doi.org/10.1016/j.jml.2009.05.004>
- Baddeley, A., & Logie, R. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44, 568-591. <https://doi.org/10.1006/jmla.2000.2756>
- Balilah, A. M. (2017). *Linguistic and cognitive measures in Arabic-speaking English language learners (ELLs) and monolingual children with and without developmental language disorder (DLD)* [Electronic thesis and dissertation repository, The University of Western

- Ontario]. <https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=7035&context=etd>
- Banerjee, A., & Davé, R. N. (2004). Validating clusters using the Hopkins statistic In *Proceedings of the IEEE International Conference on Fuzzy Systems*, Piscataway, NJ, 1, 149-153. <https://doi.org/10.1109/FUZZY.2004.1375706>
- Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child development*, 83(2), 413-422. <https://doi.org/10.1111/j.1467-8624.2011.01707.x>
- Barlow J. A. (2001). The structure of /s/-sequences: evidence from a disordered system. *Journal of Child Language*, 28(2), 291-324. <https://doi.org/10.1017/s0305000901004652>
- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37-46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Barrouillet, P., & Camos, V. (2007). Le développement de la mémoire de travail. In J. Lautrey (Ed.), *Psychologie du développement et de l'éducation* (pp. 51-86). Paris: Presses Universitaires.
- Barrouillet, P., Portrat, S., & Camos, V. (2008). On the law relating processing to storage in working memory. *Psychological Review*, 34, 1561-1564. <https://doi.org/10.1037/a0013356>
- Bar-Shalom, E., & Snyder, W. (1998). Root infinitives in child Russian: A comparison with Italian and Polish. In R. Shillcock, A. Sorace & C. Heycock (Eds.), *Language acquisition: Knowledge representation and processing. Proceedings of GALA 1997* (pp. 22-27). Edinburgh: University of Edinburgh.
- Bartke, S. (1994). *Dissociations in SLI children's inflectional morphology: New evidence from agreement inflections and noun plurals in German*. Paper presented at the meeting of the European Group for Child Language Disorders, Garderen, The Netherlands.
- Bartke, S. (1998). *Experimentelle Studien zur Flexion und Wortbildung. Pluralmorphologie und lexikalische Komposition im unauffälligen Spracherwerb und im Dysgrammatismus*. Tübingen: Niemeyer. <https://doi.org/10.1017/S1040820700002602>
- Bassiouney, R. (2009). *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press. <https://doi.org/10.1007/s10993-010-9169-0>
- Bayram, F., Rothman, J., Iverson, M., Kupisch, T., Miller, D., Puig-Mayenco, E., & Westergaard, M. (2019). Differences in use without deficiencies in competence: passives in the Turkish and German of Turkish heritage speakers in Germany. *International Journal of Bilingual Education and Bilingualism*, 22(8), 919-939. <https://doi.org/10.1080/13670050.2017.1324403>
- Bavin, E., Wilson, P., Maruff, P., & Sleeman, F. (2005). Spatio-visual memory of children with specific language impairment: Evidence for generalized processing problems. *International Journal of Language and Communication Disorders*, 40(3), 319-332. <https://doi.org/10.1080/13682820400027750>
- Beauchamp, M. L. H., & MacLeod, A. A. N. (2022). The role of exposure on school-aged bilingual language abilities—It depends on what you measure. *International Journal of Bilingual Education and Bilingualism*, 25(9), 3404-3419. <https://doi.org/10.1080/13670050.2022.2062217>
- Becker, B. (2010). The transfer of cultural knowledge in the early childhood: social and ethnic disparities and the mediating role of familial activities. *European Sociological Review*, 26(1), 17-29. <https://www.jstor.org/stable/40602475>
- Becker, R., & Beck, M. (2011). Migration, Sprachförderung und soziale Integration. Eine Evaluation der Sprachförderung von Berliner Schulkindern mit Migrationshintergrund anhand von ELEMENTPanel-Daten. In R. Becker (Ed.), *Integration durch Bildung. Bildungserwerb von jungen Migranten in Deutschland* (pp. 121-138). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-93232-3_6
- Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 41(5), 1185-1192. <https://doi.org/10.1044/jslhr.4105.1185>
- Bedore, L., & Leonard, L. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44(4), 905-924.

[https://doi.org/10.1044/1092-4388\(2001/072\)](https://doi.org/10.1044/1092-4388(2001/072))

- Bedore, L.M., & Peña, E.D. (2008). Assessment of Bilingual Children for Identification of Language Impairment: Current Findings and Implications for Practice. *International Journal of Bilingual Education and Bilingualism*, 11, 1 - 29. <https://doi.org/10.2167/beb392.0>
- Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Speech, Language, Hearing Services in Schools*, 36(3), 188-200. [https://doi.org/10.1044/0161-1461\(2005/020\)](https://doi.org/10.1044/0161-1461(2005/020))
- Bedore, L., Peña, E., Griffin, Z., & Hixon, J. G. (2016). Effects of Age of English Exposure, Current Input/Output, and grade on bilingual language performance. *Journal of Child Language*, 43(3), 687-706. <https://doi.org/10.1017/S0305000915000811>
- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., ... Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, 15(3), 616-629. <https://doi.org/10.1017/S1366728912000090>
- Beitchman, J., Jiang, H., Koyama, E., Johnson, C. J., Escobar, M., Atkinson, L., Brownlie, E. B., & Vida, R. (2008). Models and determinants of vocabulary growth from kindergarten to adulthood. *The Journal of Child Psychology and Psychiatry*, 49(6), 626-634. <https://doi.org/10.1111/j.1469-7610.2008.01878.x>
- Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., & Lancee, W. (1996). Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(6), 804-814. <https://doi.org/10.1097/00004583-199606000-00021>
- Belletti, A. (2014). Notes on Passive Object Relatives. In P. Svenonius (Ed.), *Functional Structure from Top to Toe: The 31 Cartography of Syntactic Structures*, vol 9 (pp. 97-114). Oxford: Oxford University Press.
- Belletti, A., & Contemori, C. (2010). Intervention and Attraction. On the Production of Subject and Object Relatives by Italian (Young) Children and Adults. In J. Costa, A. Castro, M. Lobo, & F. Pratas (Eds.), *Language Acquisition and Development: Proceedings of GALA 2009* (pp. 39-52). Newcastle, UK: Cambridge Scholars Publishing.
- Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*, 122(10), 1053-1069. <https://doi.org/10.1016/j.lingua.2012.02.007>
- Belletti, A., & Guasti, M. T. (2015). *The acquisition of Italian. Morphosyntax and its interfaces in different modes of acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/lald.57>
- Belletti, A., & Hamann, C. (2004). On the L2/bilingual acquisition of French by two young children with different source languages. In P. Prévost & J. Paradis (Eds.), *The acquisition of French in different contexts: Focus on functional categories* (pp. 147-174). Amsterdam: John Benjamins. <https://doi.org/10.1075/lald.32.09bel>
- Belletti, A., & Rizzi, L. (2012). Ways of Avoiding Intervention: Some Thoughts on the Development of Object Relatives, Passive, and Control. In M. Piattelli-Palmarini & R. C. Berwick (Eds.), *Rich Languages from Poor Inputs* (pp. 115-126). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199590339.003.0008>
- Belletti, A., & Rizzi, L. (2013). Intervention in grammar and processing. In I. Caponigro & C. Cecchetto (Eds.), *From Grammar to Meaning: The Spontaneous Logicality of Language* (pp. 294-311). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139519328.015>
- Benasich, A. A., Choudhury, N. A., Realpe-Bonilla, T., & Roesler, C. P. (2014). Plasticity in developing brain: active auditory exposure impacts prelinguistic acoustic mapping. *The Journal of Neuroscience: the official journal of the Society for Neuroscience*, 34(40), 13349-13363. <https://doi.org/10.1523/JNEUROSCI.0972-14.2014>
- Benasich, A. A., Gou, Z., Choudhury, N., & Harris, K. D. (2008). Early cognitive and language skills are linked to resting frontal gamma power across the first 3 years. *Behavioural brain research*, 195(2), 215-

222. <https://doi.org/10.1016/j.bbr.2008.08.049>
- Benmamoun, E. (2000). *The featural structure of functional categories: A comparative study of Arabic dialects*. Oxford: Oxford University Press
- Benmamoun, E., Albirini, A., Montrul, S., & Saadah, E. (2014). Arabic plurals and root and pattern morphology in Palestinian and Egyptian heritage speakers. *Linguistic Approaches to Bilingualism*, 4(1), 89-123. <https://doi.org/10.1075/lab.4.1.04ben>
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3-4), 129-181. <https://doi.org/10.1515/tl-2013-0009>
- Benner, S. M., & Grim, J. (2012). *Assessment of young children with special needs: A context-based approach*. Routledge. <https://doi.org/10.4324/9780203839072>
- Bentea, A. (2012). Subject vs. object relatives: what can Romanian children tell us about their acquisition? *Revue Roumaine de Linguistique*, 57(2), 203-218.
- Bentea, A., & Durrleman, S. (2014). To move or not to move? Movement, intervention and optionality in the acquisition of A'-dependencies in French. In J. Costa, A. Fiéis, M. J. Freitas, M. Lobo & A. L. Santos (Eds.), *Selected Proceedings of the Romance Turn V* (pp. 78-100). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Bentea, A., Durrleman, S., & Rizzi, L. (2016). Refining intervention: the acquisition of featural relations in object A-bar dependencies. *Lingua*, 169, 21-41. <https://doi.org/10.1016/j.lingua.2015.10.001>
- Berman, R. A. (2004). Between emergence and mastery: The long development route of language acquisition. In R. A. Berman (Ed.), *Language Development across Childhood and Adolescence, TILAR* (pp. 9-34). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.3.05ber>
- Bernardini, P., & Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: The Ivy Hypothesis. *Bilingualism: Language and Cognition*, 7(1), 49-69. <https://doi.org/10.1017/S1366728904001270>
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, 81(6), 1641-1660. <https://doi.org/10.1111/j.1467-8624.2010.01499.x>
- Bettge, S., & Oberwöhrmann, S. (2018). *Grundausswertung der Einschulungsdaten in Berlin 2017. Berlin: Senatsverwaltung für Gesundheit und Soziales Referat Gesundheitsberichterstattung, Epidemiologie, Gesundheitsinformation, Statistikstelle.* https://www.berlin.de/sen/gesundheits/_assets/service/publikationen/gesundheitsberichterstattung/veroeffentlichungen/grundausswertungen/2019-11-07-grundausswertung_esu_2017_bf.pdf
- Bhatia, T. K., & Ritchie, W. C. (Eds.). (2013). *The handbook of bilingualism and multilingualism* (2nd ed.). Malden, MA: Wiley-Blackwell.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, 13(2), 116-137. <https://doi.org/10.1191/0267658976776702>
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, 70(3), 636-644. <https://doi.org/10.1111/1467-8624.00046>
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12(1), 3-11. <https://doi.org/10.1017/S1366728908003477>
- Bialystok, E. (2011). Coordination of Executive Functions in Monolingual and Bilingual Children. *Journal of Experimental Child Psychology*, 110, 461-468. <https://doi.org/10.1016/j.jecp.2011.05.005>
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233-262. <https://doi.org/10.1037/bul0000099>
- Bialystok, E. (2018). Bilingualism and executive function: What's the connection? In D. Miller, F. Bayram, J. Rothman & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 283-305). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.54.13bia>
- Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010a). Word Mapping and Executive Functioning in Young Monolingual and Bilingual Children. *Journal of Cognition and*

- Development*, 11(4), 485-508. <https://doi.org/10.1080/15248372.2010.516420>
- Bialystok, E., Craik, F. I., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45(2), 459-464. <https://doi.org/10.1016/j.neuropsychologia.2006.10.009>
- Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual Minds. *Psychological Science in the Public Interest: a journal of the American Psychological Society*, 10(3), 89-129. <https://doi.org/10.1177/1529100610387084>
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 859-873. <https://doi.org/10.1037/0278-7393.34.4.859>
- Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240-250. <https://doi.org/10.1016/j.tics.2012.03.001>
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525-531. <https://doi.org/10.1017/S1366728909990423>
- Bialystok, E., & Martin, M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, 7(3), 325-339. <https://doi.org/10.1111/j.1467-7687.2004.00351.x>
- Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied psycholinguistics*, 35(1), 177-191. <https://doi.org/10.1017/S0142716412000288>
- Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science*, 8(6), 595-604. <https://doi.org/10.1111/j.1467-7687.2005.00451.x>
- Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, 112(3), 494-500. <https://doi.org/10.1016/j.cognition.2009.06.014>
- Bianchi, G. (2013). Gender in Italian-German Bilinguals: A comparison with German L2 learners of Italian. *Bilingualism, Language and Cognition*, 16(3), 538-557. <https://doi.org/10.1017/S1366728911000745>
- Biran, M., & Ruigendijk, E. (2015). Do case and gender information assist sentence comprehension and repetition for German-and Hebrew-speaking children? *Lingua*, 164, 215-238. <https://doi.org/10.1016/j.lingua.2015.06.012>
- Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, 56, 9-49. <https://doi.org/10.1111/j.1467-9922.2006.00353.x>
- Birdsong D. (2018). Plasticity, Variability and Age in Second Language Acquisition and Bilingualism. *Frontiers in Psychology*, 9, 81. <https://doi.org/10.3389/fpsyg.2018.00081>
- Birdsong, D., & Amengual, M. (2024). Language Dominance Effects in the Phonetics and Phonology of Bilinguals. In M. Amengual (Ed.), *The Cambridge Handbook of Bilingual Phonetics and Phonology* (pp. 655-676). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009105767.030>
- Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Psychology Press/Erlbaum (UK) Taylor & Francis. <https://doi.org/10.4324/9781315804699>
- Bishop, D. V. M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In D.V.M. Bishop & L. Leonard (Eds.), *Speech and Language Impairments in Children: Causes, Characteristics, Intervention and Outcome* (pp. 99-113). Psychology Press. <https://doi.org/10.4324/9781315784878>
- Bishop D. V. M. (2002). The role of genes in the etiology of specific language impairment. *Journal of Communication Disorders*, 35(4), 311-328. [https://doi.org/10.1016/s0021-9924\(02\)00087-4](https://doi.org/10.1016/s0021-9924(02)00087-4)
- Bishop, D. V. M. (2003). *Test for Reception of Grammar, Version 2 (TROG-2)*. London: The Psychological Corporation.

- Bishop, D. V. M. (2006a). Developmental cognitive genetics: how psychology can inform genetics and vice versa. *Quarterly Journal of Experimental Psychology* (2006), 59(7), 1153-1168. <https://doi.org/10.1080/17470210500489372>
- Bishop, D. V. M. (2006b). What Causes Specific Language Impairment in Children?. *Current directions in psychological science*, 15(5), 217-221. <https://doi.org/10.1111/j.1467-8721.2006.00439.x>
- Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, 52(6), 671-680. <https://doi.org/10.1111/1460-6984.12335>
- Bishop, D. V., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes, brain, and behavior*, 5(2), 158-169. <https://doi.org/10.1111/j.1601-183X.2005.00148.x>
- Bishop, D. V., Holt, G., Line, E., McDonald, D., McDonald, S., & Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: a longitudinal study of precursors of specific language impairment. *Journal of Neurodevelopmental Disorders*, 4(1), 3. <https://doi.org/10.1186/1866-1955-4-3>
- Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders*, 44(5), 600-615. <https://doi.org/10.1080/13682820802259662>
- Bishop, D. V. M., & Norbury, C. F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: a study using standardized diagnostic instruments. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 43, 917-929. <https://doi.org/10.1111/1469-7610.00114>
- Bishop, D. V., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: evidence from a twin study. *Developmental medicine and child neurology*, 37(1), 56-71. <https://doi.org/10.1111/j.1469-8749.1995.tb11932.x>
- Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *Journal of child psychology and psychiatry, and allied disciplines*, 37(4), 391-403. <https://doi.org/10.1111/j.1469-7610.1996.tb01420.x>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of child psychology and psychiatry, and allied disciplines*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
- Black, M., & Chiat, S. (2003). Noun-verb dissociations: A multi-faceted phenomenon. *Journal of Neurolinguistics*, 16(2), 231-250. [https://doi.org/10.1016/S0911-6044\(02\)00017-9](https://doi.org/10.1016/S0911-6044(02)00017-9)
- Blishen, B. R., Carroll, W. K., & Moore, C. (1987). The 1981 socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology*, 24, 465-488. <https://doi.org/10.1111/j.1755-618X.1987.tb00639.x>
- Blom, E. (2003). *From root infinitives to finite sentences: The acquisition of verbal inflections and auxiliaries* [Doctoral dissertation, University of Utrecht, The Netherlands].
- Blom, E. (2019). Domain-general cognitive ability predicts bilingual children's receptive vocabulary in the majority language. *Language Learning*, 69(2), 292-322. <https://doi.org/10.1111/lang.12333>
- Blom, E., & Baayen, H. R. (2013). The impact of verb form, sentence position, home language, and second language proficiency on subject-verb agreement in child second language Dutch. *Applied Psycholinguistics*, 34(4), 777-811. <https://doi.org/10.1017/S0142716412000021>
- Blom, E., Berke, R., Shaya, N., & Adi-Japha, E. (2021a). Cognitive flexibility in children with Developmental Language Disorder: Drawing of nonexistent objects. *Journal of Communication Disorders*, 93, 106137. <https://doi.org/10.1016/j.jcomdis.2021.106137>
- Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure?. *Research in Developmental Disabilities*, 55, 301-311. <https://doi.org/10.1016/j.ridd.2016.05.001>
- Blom, E., & Boerma, T. (2017). Effects of language impairment and bilingualism across domains:

- Vocabulary, morphology and verbal memory. *Linguistic Approaches to Bilingualism*, 7(3-4), 277-300. <https://doi.org/10.1075/lab.15018.blo>
- Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts. *Frontiers in Psychology*, 8, 552. <https://doi.org/10.3389/fpsyg.2017.00552>
- Blom, E., Boerma, T., & de Jong, J. (2019). First language attrition and developmental language disorder. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 108-120). Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780198793595.013.10>
- Blom, E., Boerma, T., Karaca, F., de Jong, J., & Küntay, A. C. (2022). Grammatical development in both languages of bilingual Turkish-Dutch children with and without developmental language disorder. *Frontiers in Communication*, 7, 1059427. <https://doi.org/10.3389/fcomm.2022.1059427>
- Blom, E., & Bosma, E. (2016). The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian-Dutch children. *Journal of Child Language*, 43(3), 581-607. <https://doi.org/10.1017/S0305000915000574>
- Blom, E., Bosma, E., & Heeringa, W. (2021b). Regular and irregular inflection in different groups of bilingual children and the role of verbal short-term and verbal working memory. *Languages*, 6(1), 1-23. <https://doi.org/10.3390/languages6010056>
- Blom, E., de Jong, J., Orgassa, A., Baker, A., & Weerman, F. (2013). Verb inflection in monolingual Dutch and sequential bilingual Turkish-Dutch children with and without SLI. *International Journal of Language & Communication Disorders*, 48(4), 382-393. <https://doi.org/10.1111/1460-6984.12013>
- Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, 128, 105-119. <https://doi.org/10.1016/j.jecp.2014.06.007>
- Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 56(1), 281-294. [https://doi.org/10.1044/1092-4388\(2012/11-0112\)](https://doi.org/10.1044/1092-4388(2012/11-0112))
- Blom, E., & Paradis, J. (2015). Sources of individual differences in the acquisition of tense inflection by English second language learners with and without specific language impairment. *Applied Psycholinguistics*, 36(4), 953-976. <https://doi.org/10.1017/S014271641300057X>
- Blom, E., Paradis, J., & Sorenson Duncan, T. (2012). Effects of input properties, vocabulary size and L1 on the development of third person singular -s in child L2 English. *Language Learning*, 62(3), 965-994. <https://doi.org/10.1111/j.1467-9922.2012.00715.x>
- Blom, E., Polišenská, D., & Weerman, F. (2008). Articles, adjectives and age of onset: the acquisition of Dutch grammatical gender. *Second Language Research*, 24(3), 297-331. <https://doi.org/10.1177/0267658308090183>
- Blom, E., Soto-Corominas, A., Attar, Z., Daskalaki, E., & Paradis, J. (2021c). Interdependence between L1 and L2: The case of Syrian children with refugee backgrounds in Canada and the Netherlands. *Applied Psycholinguistics*, 42(5), 1159-1194. <https://doi.org/10.1017/S0142716421000229>
- Bloom, L., Rispoli, M., Gartner, B., & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language*, 16(1), 101-120. <https://doi.org/10.1017/s0305000900013465>
- Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible?. *Journal of Communication Disorders*, 66, 65-76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research*, 58(6), 1747-1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058
- Boerma, T., Leseman, P., Wijnen, F., & Blom, E. (2017). Language Proficiency and Sustained Attention in Monolingual and Bilingual Children with and without Language Impairment. *Frontiers in psychology*, 8, 1241. <https://doi.org/10.3389/fpsyg.2017.01241>

- Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (Version 5.3.56) <http://www.praat.org/>
- Bohman, T.M., Bedore, L.M., Peña, E.D., Mendez-Perez, A., & Gillam, R.B. (2010). What you hear and what you say: language performance in Spanish-English bilinguals. *International Journal of Bilingual Education and Bilingualism*, 13(3), 325-344. <http://dx.doi.org/10.1080/13670050903342019>
- Bohlmann, N. L., Maier, M. F., & Palacios, N. (2015). Bidirectionality in self-regulation and expressive vocabulary: Comparisons between monolingual and dual language learners in preschool. *Child Development*, 86(4), 1094-1111. <https://doi.org/10.1111/cdev.12375>
- Bonifacci, P., Atti, E., Casamenti, M., Piani, B., Porrelli, M., & Mari, R. (2020). Which Measures Better Discriminate Language Minority Bilingual Children with and without Developmental Language Disorder? A Study Testing a Combined Protocol of First and Second Language Assessment. *Journal of Speech, Language, and Hearing Research*, 63(6), 1898-1915. https://doi.org/10.1044/2020_JSLHR-19-00100
- Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science*, 14(2), 256-269. <https://doi.org/10.1111/j.1467-7687.2010.00974.x>
- Bonnesen, M. (2009). The status of the “weaker” language in unbalanced French/German bilingual language acquisition. *Bilingualism: Language and Cognition*, 12(2), 177-192. <https://doi.org/10.1017/S1366728908003982>
- Bortolini, U., & Leonard, L. B. (2000). Phonology and children with specific language impairment: status of structural constraints in two languages. *Journal of Communication Disorders*, 33(2), 131-150. [https://doi.org/10.1016/s0021-9924\(99\)00028-3](https://doi.org/10.1016/s0021-9924(99)00028-3)
- Bos, P. (1997). *Development of bilingualism: A study of school-age Moroccan children in the Netherlands*. Tilburg University Press.
- Bosch, L. (2024). Acquisition of Segmental Phonology in Child Bilingualism. In M. Amengual (Ed.), *The Cambridge Handbook of Bilingual Phonetics and Phonology* (pp. 339-358). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009105767.016>
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., ... & Cohen, J. F. (2015). STARD 2015. An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Clinical Chemistry*, 61(12), 1446-1452. <https://doi.org/10.1373/clinchem.2015.246280>
- Botting, N. (2005). Non-verbal cognitive development and language impairment. *The Journal of Child Psychology and Psychiatry*, 46(3), 317-326. <https://doi.org/10.1111/j.1469-7610.2004.00355.x>
- Botting N. (2020). Language, literacy and cognitive skills of young adults with developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, 55(2), 255-265. <https://doi.org/10.1111/1460-6984.12518>
- Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language and Communication Disorders*, 36(4), 421-432. <https://doi.org/10.1080/13682820110074971>
- Botting, N., & Conti-Ramsden, G. (2003). Autism, primary pragmatic difficulties, and specific language impairment: can we distinguish them using psycholinguistic markers?. *Developmental medicine and child neurology*, 45(8), 515-524. <https://doi.org/10.1017/s0012162203000963>
- Botting, N., Durkin, K., Toseeb, U., Pickles, A., & Conti-Ramsden, G. (2016a). Emotional health, support, and self-efficacy in young adults with a history of language impairment. *British Journal of Developmental Psychology*, 34(4), 538-554. <https://doi.org/10.1111/bjdp.12148>
- Botting, N., & Marshall, C. (2017). Domain-specific and Domain-general Approaches to Developmental Disorders. In L.C. Centifanti & D.M. Williams (Eds.), *The Wiley Handbook of Developmental Psychopathology* (pp. 139-159). <https://doi.org/10.1002/9781118554470.ch6>
- Botting, N., Toseeb, U., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2016b). Depression and anxiety change from adolescence to adulthood in individuals with and without language impairment. *PLoS*

- ONE, 11(7), e0156678. <https://doi.org/10.1371/journal.pone.0156678>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652. <https://doi.org/10.1037/0033-295X.108.3.624>
- Botwinik, I., Bshara, R., & Armon-Lotem, S. (2015). Children's production of relative clauses in Palestinian Arabic: Unique errors and their movement account. *Lingua*, 156, 40-56. <https://doi.org/10.1016/j.lingua.2014.10.007>
- Boyle, W., Lindell, A.K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning*, 63(2), 211-242. <https://doi.org/10.1111/lang.12003>
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371-99. <https://doi.org/10.1146/annurev.psych.53.100901.135233>
- Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. *American Journal of Speech-Language Pathology*, 24(2), 126-138. https://doi.org/10.1044/2015_AJSLP-13-0106
- Brandt, S., Kidd, E., Lieven, E., & Tomasello, M. (2009). The discourse bases of relativization: An investigation of young German and English-speaking children's comprehension of relative clauses. *Cognitive Linguistics*, 20(3), 539-570. <https://doi.org/10.1515/COGL.2009.024>
- Breitfeld, E., Potter, C. E., & Lew-Williams, C. (2021). Children simultaneously learn multiple dimensions of information during shared book reading. *Journal of Cognition and Development: official journal of the Cognitive Development Society*, 22(5), 744-766. <https://doi.org/10.1080/15248372.2021.1939353>
- Bridges, K., & Hoff, E. (2014). Older sibling influences on the language environment and language development of toddlers in bilingual homes. *Applied psycholinguistics*, 35(2), 225-241. <https://doi.org/10.1017/S0142716412000379>
- Briscoe, J., Bishop, D. V., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry*, 42(3), 329-340. <https://doi.org/10.1111/1469-7610.00726>
- Briscoe, J., & Rankin, P. M. (2007). Exploration of a 'double-jeopardy' hypothesis within WM profiles for children with specific language impairment. *International Journal of Language of Communication Disorders*, 44(2), 236-250. <https://doi.org/10.1080/13682820802028760>
- Brito, N. H., Fifer, W. P., Myers, M. M., Elliott, A. J., & Noble, K. G. (2016). Associations among family socioeconomic status, EEG power at birth, and cognitive skills during infancy. *Developmental cognitive neuroscience*, 19, 144-151. <https://doi.org/10.1016/j.dcn.2016.03.004>
- Brito, N., & Noble, K. (2014). Socioeconomic status and structural brain development. *Frontiers in Neuroscience*, 8, 276, 1-12. <https://doi.org/10.3389/fnins.2014.00276>
- Brito, N. H., Noble, K. G., & Pediatric Imaging, Neurocognition, Genetics Study (2018). The independent and interacting effects of socioeconomic status and dual-language use on brain structure and cognition. *Developmental science*, 21(6), e12688. <https://doi.org/10.1111/desc.12688>
- Brito, N., Sebastian-Galles, N., & Barr, R. (2015). Differences in language exposure and its effects on memory flexibility in monolingual, bilingual, and trilingual infants. *Bilingualism: Language and Cognition*, 18, 670-682. <https://doi.org/10.1017/S1366728914000789>
- Brizzolara, D., Gasperini, F., Pfanner, L., Cristofani, P., Casalini, C., & Chilosi, A. M. (2011). Long-term reading and spelling outcome in Italian adolescents with a history of specific language impairment. *Cortex*, 47(8), 955-973. <https://doi.org/10.1016/j.cortex.2011.02.009>
- Brownlie, E. B., Bao, L., & Beitchman, J. (2016). Childhood Language Disorder and Social Anxiety in Early Adulthood. *Journal of Abnormal Child Psychology*, 44(6), 1061-1070. <https://doi.org/10.1007/s10802-015-0097-5>
- Brownlie, E., Graham, E., Bao, L., Koyama, E., & Beitchman, J. (2017). Language disorder and retrospectively reported sexual abuse of girls: severity and disclosure. *Journal of child psychology and*

- psychiatry, 58(10), 1114-1121. <https://doi.org/10.1111/jcpp.12723>
- Brun, D., Avrutin, S., & Babyonyshev, M. (1999). Aspect and its temporal interpretation during the optional infinitive stage in Russian. In A. Greenhill, H. Littlefield & C. Tano, *Proceedings of BUCLD 23* (pp. 120-131). Somerville, MA: Cascadilla Press.
- Bruns-Philipps, E., Hespe-Jungesblut, K., Jahn, N., Schicktan, C., & Zühlke C. (2018). *Kindergesundheit im Einschulungsalter: Ergebnisse der Schuleingangsuntersuchung 2017*. Gesundheitsberichterstattung für Niedersachsen. https://www.nlga.niedersachsen.de/startseite/gesundheitsberichterstattung/gesundheitsberichte/schuleingangsuntersuchung_seu/berichte_zur_kindergesundheit_im_einschulungsalter/spezialberichte-133588.html
- Brustad, K. E. (2000). *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Georgetown: Georgetown University Press.
- Bryan, K., Garvani, G., Gregory, J., & Kilner, K. (2015). Language difficulties and criminal justice: the need for earlier identification. *International Journal of Language and Communication Disorders*, 50, 763-775. <https://doi.org/10.1111/1460-6984.12183>
- Bühning, D. (1995). *The 59th Bridge accent: the meaning of focus and topic*. [Doctoral dissertation, University of Köln].
- Bulheller, S., & Häcker, H. (2002). *Coloured progressive matrices—Deutsche Bearbeitung und Normierung* [German adaptation and norms]. Frankfurt: Swets & Zeitlinger.
- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental neuropsychology*, 33(3), 205-228. <https://doi.org/10.1080/87565640801982312>
- Bundesinstitut für Arzneimittel und Medizinprodukte (Ed.) (2020). *Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD) (10. Revision), German Modification*, Version 2021. Retrieved on 08.04.2022 <https://www.dimdi.de/static/de/klassifikationen/icd/icd-10-gm/kode-suche/htmlgm2021/block-f80-f89.htm#F80>
- Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron*, 33(2), 301-311. [https://doi.org/10.1016/s0896-6273\(01\)00583-9](https://doi.org/10.1016/s0896-6273(01)00583-9)
- Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cerebral cortex (New York, N.Y.: 1991)*, 15(3), 239-249. <https://doi.org/10.1093/cercor/bhh126>
- Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual spanish- and english-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research*, 50(4), 999-1014. [https://doi.org/10.1044/1092-4388\(2007/070\)](https://doi.org/10.1044/1092-4388(2007/070))
- Burke, H.L., & Coady, J.A. (2015). Nonword repetition errors of children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, 50, 337-346. <https://doi.org/10.1111/1460-6984.12136>
- Byers-Heinlein, K., & Werker, J.F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, 12, 815-823. <https://doi.org/10.1111/j.1467-7687.2009.00902.x>
- Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, 128, 407-416. <http://dx.doi.org/10.1016/j.cognition.2013.05.010>
- Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, 130(3), 278-288. <https://doi.org/10.1016/j.cognition.2013.11.015>
- Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 68, 759-778.

<https://doi.org/10.1080/17470218.2014.966248>

- Cañas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, 46(5), 482-501. <https://doi.org/10.1080/0014013031000061640>
- Cantone, K. F., Müller, N., Schmitz, K., & Kupisch, T. (2008). Rethinking language dominance in bilingual children. *Linguistische Berichte* 215, 307-343. https://doi.org/10.46771/2366077500215_2
- Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, 6(3), 184-205. <https://doi.org/10.1080/15475441.2010.484379>
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, 15, 17-29.
- Carlson, S., & Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11(2), 282-298. <https://doi.org/10.1111/j.1467-7687.2008.00675.x>
- Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy (Person> Number>Gender) and implications for linguistic theory. *Lingua*, 115, 259-85. <https://doi.org/10.1016/j.lingua.2003.10.006>
- Carroll, C., & Dockrell, J. (2010). Leaving special school: post-16 outcomes for young adults with specific language impairment. *European Journal of Special Needs Education*, 25(2), 131-147. <https://doi.org/10.1080/08856251003658660>
- Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., Roncoli, S., & Burani, C. (2007). Non-word repetition in children with specific language impairment: a deficit in phonological working memory or in long-term verbal knowledge?. *Cortex: a journal devoted to the study of the nervous system and behavior*, 43(6), 769-776. [https://doi.org/10.1016/s0010-9452\(08\)70505-7](https://doi.org/10.1016/s0010-9452(08)70505-7)
- Case, R. (1985). *Intellectual development. Birth to adulthood*. New York: Academic Press.
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R., & Coltheart, M. (2007). *PALPA-P Provas de Avaliacao da Linguagem e da Afasia em Portugues*. Lisbon: CEGOC.
- Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, 49, 649-671. <http://dx.doi.org/10.1111/1460-6984.12082>
- Catts, H. W., Fey, M. E., Weismer, S. E., & Bridges, M. S. (2014). The relationship between language and reading abilities. In J.B. Tomblin & M. A. Nippold (Eds.), *Understanding individual differences in language development across the school years* (pp. 158-179). New York: Psychology Press.
- Centanni, T. M., Green, J. R., Iuzzini-Seigel, J., Bartlett, C. W., & Hogan, T. P. (2015). Evidence for the multiple hits genetic theory for inherited language impairment: A case study. *Frontiers in Genetics*, 6, 272. <https://doi.org/10.3389/fgene.2015.00272>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61(6), 1-36. <https://doi.org/10.18637/jss.v061.i06>
- Chen, X., and Schwartz, M. (2018). Morphological awareness and literacy in second language learners: a cross-language perspective. *Reading and Writing*, 31, 1685-1694, <https://doi.org/10.1007/s11145-018-9866-1>
- Chevrie-Muller, C., & Maillart, C. (2009). L'évaluation de la morphosyntaxe chez des enfants d'âge scolaire (7-12 ans): les épreuves issues de la batterie L2MA-2. *Rééducation Orthophonique*, 238, 101-115.
- Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes*, 16(2-3), 113-142. <https://doi.org/10.1080/01690960042000012>
- Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 125-151). Bristol: Multilingual Matters. <https://doi.org/10.21832/9781783093137-008>
- Chiat, S., Armon-Lotem, S., Marinis, T., Polišenská, K., Roy, P., & Seeff-Gabriel, B. (2013). Assessment

- of language abilities in sequential bilingual children: The potential of sentence imitation tasks. In V. C. M. Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 56-89). Bristol, UK: Multilingual Matters. <https://doi.org/10.21832/9781783090105-005>
- Chiat, S., & Polišenská, K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. *Journal of Speech, Language, and Hearing Research*, 59(5), 1179-1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293
- Chien, Y. C., & Wexler, K. (1990). Children's Knowledge of Locality Conditions in Binding as Evidence for the Modularity of Syntax and Pragmatics. *Language Acquisition*, 1(3), 225-295. https://doi.org/10.1207/s15327817la0103_2
- Chierchia, G., & McConnell-Ginet, S. (1990). *Meaning and grammar: An introduction to semantics*. Cambridge, Mass.: MIT Press
- Chilla, S. (2008). *Erstsprache, Zweitsprache, Spezifische Sprachentwicklungsstörung? Eine Untersuchung der deutschen Hauptsatzstruktur durch sukzessiv-bilinguale Kinder mit türkischer Erstsprache*. Hamburg: Dr. Kovač.
- Chilla, S. (2022). Assessment of developmental language disorders in bilinguals: Immigrant Turkish as a bilingual challenge in Germany. In E. Saiegh-Haddad, L. Laks, & C. McBride (Eds.), *Handbook of Literacy in Diglossia and in Dialectal Contexts. Literacy Studies, vol 22*. Springer, Cham. https://doi.org/10.1007/978-3-030-80072-7_16
- Chilla, S., & Abed Ibrahim, L. (2026). Zum Einsatz von LITMUS-Satz- und Kunstwortwiederholungstests bei Schüler*innen mit Forced-Displacement-Biografie. *Sprache Stimme Gehör (advanced online publication)*. DOI: 10.1055/a-2736-9286
- Chilla, S., & Hamann, C. (2018). Sprachliche Heterogenität als Herausforderung und Chance für den inklusiven Fremdsprachenunterricht. *Zeitschrift Fremdsprachenforsch*, 32(1), 105-126. <https://doi.org/10.1055/s-0043-123986>
- Chilla, S., & Hamann, C. (2021). Sprachliche Heterogenität als Herausforderung und Chance für den inklusiven Fremdsprachenunterricht. *Zeitschrift für Fremdsprachenforschung*, 32(1), 105-126.
- Chilla, S., Hamann, C., Prévost, P., Abed Ibrahim, L., Ferré, S., dos Santos, C., Zebib, R., & Tuller, L. (2021). The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German. In S. Armon-Lotem & K. K. Grohmann (Eds.), *Language Impairment in Multilingual Settings: LITMUS in action across Europe, TILAR* (pp. 227-262). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.29.08chi>
- Chilla, S., & Şan, H. (2017). Möglichkeiten und Grenzen der Diagnostik ersprachlicher Fähigkeiten: Türkisch-deutsche und türkisch-französische Kinder im Vergleich. In C. Yıldız, N. Topaj., R. Thomas & I. Gülzow (Eds.), *Die Zukunft der Mehrsprachigkeit im deutschen Bildungssystem: Russisch und Türkisch im Fokus* (pp. 175-205). Berlin: Peter Lang Verlag. <https://doi.org/10.3726/b11030>
- Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106(1), 41-61. <http://doi.org/10.1016/j.jecp.2010.01.003>
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York, NY: Praeger.
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
- Chomsky, N. (2000). Minimalist inquiries. In R. Martin, D. Michaels & J. Uriagereka (Rds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp.89-155). Cambridge, MA: MIT Press.
- Chomsky, N. (2001). Derivation by Phase. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1-52). The MIT Press. <https://doi.org/10.7551/mitpress/4056.003.0004>
- Chomsky, N. (2005). Three factors in language design. *Linguistic inquiry*, 36(1), 1-22. <https://doi.org/10.1162/0024389052993655>
- Chondrogianni, V. (2008). The acquisition of the D-domain in child L2 Greek: Effects of age and construction. In B. Haznedar & R. Slabakova (Eds.), *Child L2 acquisition: A generative*

- perspective* (pp. 97-145). Amsterdam: John Benjamins.
- Chondrogianni, V. (2018). Child L2 acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 315-323). John Benjamins. <https://doi.org/10.1075/sibil.54.06cho>
- Chondrogianni, V. (2023). Individual differences differentially influence language domains and learning mechanisms. *Journal of Child Language*, 50(4), 823-826. <https://doi.org/10.1017/S0305000923000028>
- Chondrogianni, V., & Daskalaki, E. (2023). Heritage language use in the country of residence matters for language maintenance, but short visits to the homeland can boost heritage language outcomes. *Frontiers in Language Sciences*, 2, 1230408. <https://doi.org/10.3389/flang.2023.1230408>
- Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, 1(3), 318-345. <https://doi.org/10.1075/lab.1.3.05cho>
- Chondrogianni, V., & Marinis, T. (2016). L2 children do not fluctuate: Production and online processing of indefinite articles in Turkish-speaking child learners of English. In B. Haznedar & N. H. Ketrez (Eds.), *The acquisition of Turkish in childhood* (pp. 361-388). Amsterdam: John Benjamins. <https://doi.org/10.1075/tilar.20.16cho>
- Chondrogianni, V., & Roesch, A. (2012). Exhaustivity in French-German L1-T, L1-SLI, 2L1-TD and L2-TD children. Paper presented at the COST IS0804 Meeting, Berlin, Germany.
- Chondrogianni, V., & Vasić, N. (2016). Cross-linguistic aspects in child L2 acquisition. *International Journal of Bilingualism*, 20(4), 361-368. **Fehler! Linkreferenz ungültig.**
- Chondrogianni, V., & Schwartz, R. G. (2020). Case marking and word order in Greek heritage children. *Journal of Child Language*, 47(4), 766-795. <https://doi.org/10.1017/S0305000919000849>
- Choudhury, N., & Benasich, A. (2003). A family aggregation study: The influence of family history and other risk factors on language development. *Journal of Speech, Language and Hearing Research*, 46, 261-272. [https://doi.org/10.1044/1092-4388\(2003/021\)](https://doi.org/10.1044/1092-4388(2003/021))
- Chung, H.J., Weyandt, L.L., Swentosky, A. (2014). The Physiology of Executive Functioning. In S. Goldstein & J. Naglieri (Eds.), *Handbook of Executive Functioning* (13-27). New York: Springer. https://doi.org/10.1007/978-1-4614-8106-5_2
- Claessen, M., Heath, S., Fletcher, J., Hogben, J., & Leitão, S. (2009). Quality of phonological representations: a window into the lexicon?. *International Journal of Language & Communication Disorders*, 44(2), 121-144. <https://doi.org/10.1080/13682820801966317>
- Claessen, M., & Leitão, S. (2012). Phonological representations in children with SLI. *Child Language Teaching and Therapy*, 28, 211-223. <https://doi.org/10.1177/0265659012436851>
- Claessen, M., Leitão, S., Kane, R., & Williams, C. (2013). Phonological processing skills in specific language impairment. *International Journal of Speech-Language Pathology*, 15(5), 471-483. <https://doi.org/10.3109/17549507.2012.753110>
- Clahsen, H. (1986a). *Die Profilanalyse. Ein linguistisches Verfahren für die Sprachdiagnose im Vorschulalter*. Berlin: Marhold
- Clahsen, H. (1986b). Verb inflections in German child language: acquisition of agreement markings and the functions they encode. *Linguistics*, 24(1), 79-122. <https://doi.org/10.1515/ling.1986.24.1.79>
- Clahsen, H. (1988). *Normale und gestörte Kindersprache. Linguistische Untersuchungen zum Erwerb von Syntax und Morphologie*. Amsterdam: Benjamins. <https://doi.org/10.1075/z.33>
- Clahsen, H. (1989). The Grammatical Characterization of Developmental Dysphasia. *Linguistics* 27(5), 897-920. <https://doi.org/10.1515/ling.1989.27.5.897>
- Clahsen, H. (1991a). Child Language and Developmental Dysphasia: Linguistic Studies of the Acquisition of German. Amsterdam: John Benjamins. <https://doi.org/10.1075/sspcl.2>
- Clahsen, H. (1991b). Constraints on parameter setting: A grammatical analysis of some acquisition in stages in German child language. *Language Acquisition*, 1(4), 361-391. https://doi.org/10.1207/s15327817la0104_3

- Clahsen, H. (1999). Linguistic Perspectives on Specific Language Impairment. In T. Bhatia & W. Ritchie (Eds.), *Handbook of Child Language Acquisition* (pp.675-704). London: Academic Press.
- Clahsen, H., Bartke, S., & Göllner, S. (1997). Formal features in impaired grammars: A comparison of English and German SLI children. *Journal of Neurolinguistics*, 10(2-3), 151-171. [https://doi.org/10.1016/S0911-6044\(97\)00006-7](https://doi.org/10.1016/S0911-6044(97)00006-7)
- Clahsen, H., & Dalalakis, J. (1999). *Tense and agreement in Greek SLI: A case study*. Colchester, UK: Department of Language and Linguistics, University of Essex.
- Clahsen, H., Kursawe, C., Penke, M. (1996a). Introducing CP: Wh-questions and Subordinate Clauses in German Child Language. *Essex Research Reports in Linguistics*, 7, 1-28.
- Clahsen, H., Marcus, G., Bartke, S., & Wiese, R. (1996b). Compounding and inflection in German child language. In G. Booji & J. Van Marle (Eds.), *Yearbook of morphology* 1995 (pp. 115-142). Dordrecht: Springer. https://doi.org/10.1007/978-94-017-3716-6_7
- Clahsen, H., & Penke, M. (1992). The acquisition of agreement morphology and its syntactic consequences. In Meisel, J.M. (Ed.), *The acquisition of verb placement. Functional categories and V2 phenomena in language acquisition* (pp.181-223). Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-2803-2_7
- Clahsen, H., Penke, M., & Parodi, T. (1993). Functional Categories in Early Child German. *Language Acquisition*, 3(4), 395-429. <http://www.jstor.org/stable/20011408>
- Clahsen, H., & Rothweiler, M. (1993). Inflectional rules in children's grammars: Evidence from German participles. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 1-34). Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-3710-4_1
- Clahsen, H., Rothweiler, M., Sterner, F., & Chilla, S. (2014). Linguistic markers of specific language impairment in bilingual children: The case of verb morphology. *Clinical Linguistics & Phonetics*, 28(9), 709-721. <https://doi.org/10.3109/02699206.2014.886726>
- Cleave, P. L., & Rice, M. L. (1997). An examination of the morpheme BE in children with specific language impairment: the role of contractibility and grammatical form class. *Journal of Speech, Language, and Hearing Research*, 40(3), 480-492. <https://doi.org/10.1044/jslhr.4003.480>
- Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders--a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 46(2), 128-149. <https://doi.org/10.1111/j.1469-7610.2004.00342.x>
- Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, 43(1), 1-40. <https://doi.org/10.1080/13682820601116485>
- Coady, J., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders*, 45(4), 494-509. <https://doi.org/10.3109/13682820903222783>
- Cobo-Lewis, A. B., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002). Effects of bilingualism and bilingual education on oral and written Spanish skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 64-97). Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781853595721-006>
- Coene, M., & Avram, L. (2011). An asymmetry in the acquisition of accusative clitics in child Romanian. *Studies on Language Acquisition*, 43, 39-68. <https://doi.org/10.1515/9783110259179.39>
- Cohen J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Collins, C. (2005). A smuggling approach to the passive in English. *Syntax*, 8(2), 81-120. <https://doi.org/10.1111/j.1467-9612.2005.00076.x>
- Condon, S., & Régnard, C. (2010). Héritage et pratiques linguistiques des descendants d'immigrés en France. *Hommes et Migrations*, 1288, 44-56. <https://doi.org/10.4000/hommesmigrations.854>
- Conlin, J. A., & Gathercole, S. E. (2006). Lexicality and interference in working memory in children and adults. *Journal of Memory and Language*, 55, 363-380. <https://doi.org/10.1016/j.jml.2006.05.003>

- Conroy, A., Takahashi, E., Lidz, J., & Phillips, C. (2009). Equal Treatment for All Antecedents: How Children Succeed with Principle B. *Linguistic Inquiry*, 40(3), 446-486. <https://doi.org/10.1162/ling.2009.40.3.446>
- Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, 46(5), 1029-1037. [https://doi.org/10.1044/1092-4388\(2003/082\)](https://doi.org/10.1044/1092-4388(2003/082))
- Conti-Ramsden, G., & Botting, N. (1999). Classification of children with specific language impairment: Longitudinal considerations. *Journal of Speech, Language, and Hearing Research*, 42(5), 1195-1204. <https://doi.org/10.1044/jslhr.4205.1195>
- Conti-Ramsden, G., & Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, 49(5), 516-525. <https://doi.org/10.1111/j.1469-7610.2007.01858.x>
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001a). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry*, 42, 741-748. <https://doi.org/10.1111/1469-7610.00770>
- Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001b). Follow-up of children attending infant language units: outcomes at 11 years of age. *International Journal of Language & Communication Disorders*, 36(2), 207-219. <https://doi.org/10.1080/13682820121213>
- Conti-Ramsden, G., & Durkin, K. (2011). Specific language impairment. In D. Skuse, H. Bruce, L. Dowdney, & D. Mrazek (Eds.), *Child psychology and psychiatry: frameworks for practice* (pp. 180-186). Chichester: Wiley-Blackwell. <https://doi.org/10.1002/9781119993971.ch29>
- Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology review*, 22(4), 384-401. <https://doi.org/10.1007/s11065-012-9208-z>
- Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, 53(2), 237-255. <https://doi.org/10.1111/1460-6984.12338>
- Conti-Ramsden, G., St Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: from childhood to adolescence. *Journal of Speech, Language, and Hearing Research*, 55(6), 1716-1735. [https://doi.org/10.1044/1092-4388\(2012/10-0182\)](https://doi.org/10.1044/1092-4388(2012/10-0182))
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, 12(5), 769-786. <https://doi.org/10.3758/bf03196772>
- Cooperson, S. J., Bedore, L. M., & Peña, E. D. (2013). The relationship of phonological skills to language skills in Spanish-English-speaking bilingual children. *Clinical Linguistics & Phonetics*, 27(5), 371-389. <https://doi.org/10.3109/02699206.2013.782568>
- Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish-English bilinguals from 22 to 30 months: implications for assessment. *Journal of Speech, Language, and Hearing Research*, 56(5), 1637-1649. [https://doi.org/10.1044/1092-4388\(2013/11-0044\)](https://doi.org/10.1044/1092-4388(2013/11-0044))
- Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 267-295. <https://doi.org/10.1177/0267658308090182>
- Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, 24(3), 183-203. <https://doi.org/10.1007/BF02145355>
- Correia, L., & Flores, C. (2017). The Role of Input Factors in the Lexical Development of European Portuguese as a Heritage Language in Portuguese-German Bilingual Speakers. *Languages*, 2(4), 30. <https://doi.org/10.3390/languages2040030>
- Correia, L., Lobo, M., & Flores, C. (2024). Sentence repetition task for European Portuguese: Results from a study with monolingual and Portuguese-German bilingual children. *Language Acquisition*, 32(4), 453-483. <https://doi.org/10.1080/10489223.2024.2346586>
- Costa, J., Friedmann, N., Silva, C., & Yachini, M. (2014). The boy that the chef cooked: Acquisition of PP

- relatives in European Portuguese and Hebrew. *Lingua*, 150, 386-409. <https://doi.org/10.1016/j.lingua.2014.08.005>
- Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, 113(2), 135-149. <https://doi.org/10.1016/j.cognition.2009.08.001>
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition*, 106(1), 59-86. <https://doi.org/10.1016/j.cognition.2006.12.013>
- Courage, M. L., & Cowan, N. (Eds.). (2022). *The development of memory in infancy and childhood*. Psychology Press. <https://doi.org/10.4324/9781003016533>
- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, 51(1), 42-100. <https://doi.org/10.1016/j.cogpsych.2004.12.001>
- Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: what does it take to make them work?. *Psychological review*, 119(3), 480-499. <https://doi.org/10.1037/a0027791>
- Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General*, 143(5), 1806-1836. <https://doi.org/10.1037/a0036814>
- Crago, M., & Paradis, J. (2003). Two of a kind? Commonalities and variation in languages and language learners. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Towards a definition of specific language impairment* (pp. 97-110). Mahwah, NJ: Erlbaum. <https://doi.org/10.1017/S0305000904216671>
- Crain, S., & Fodor, J.D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen & A. Zwicky, (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 94-128). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511597855.004>
- Crutchley, A., Botting, N., & Conti-Ramsden, G. (1997). Bilingualism and specific language impairment in children attending language units. *European Journal of Disorders of Communication: the journal of the College of Speech and Language Therapists*, London, 32(2), 267-276. <https://doi.org/10.3109/13682829709020409>
- Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research*, 49(2), 222-251. <https://doi.org/10.3102/003465430490022>
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters.
- Cummins, J. (2005). A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal*, 89(4), 585-592. <https://www.jstor.org/stable/3588628>
- Cummins, J. (2017). Teaching for transfer in multilingual school contexts. In: García O., Lin A., & May, S. (Eds.) *Bilingual and Multilingual Education. Encyclopedia of Language and Education* (3rd ed.) (pp. 103- 115). New York, NY: Springer International Publishing. https://doi.org/10.1007/978-3-319-02258-1_8
- Czapka, S., Topaj, N., & Gagarina, N. (2021). A Four-Year Longitudinal Comparative Study on the Lexicon Development of Russian and Turkish Heritage Speakers in Germany. *Languages*, 6(1), 27. <https://doi.org/10.3390/languages6010027>
- Cztinglar, C., Korecky-Kröll, K., Uzunkaya-Sharma, K., & Dressler, W. (2015). Wie beeinflusst der sozioökonomische Status den Erwerb der Erst- und Zweitsprache? In A. Ziegler & K. M. Köpcke (Eds.), *Deutsch als Zweitsprache in Schule und Unterricht* (pp. 207-240). Berlin: De Gruyter. <https://doi.org/10.1515/9783110367171-010>
- Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in neurosciences*, 38(9), 571-578. <https://doi.org/10.1016/j.tins.2015.07.003>

- Dalal, D. K., & Zickar, M. J. (2012). Some Common Myths About Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression. *Organizational Research Methods*, 15(3), 339-362. <https://doi.org/10.1177/1094428111430>
- Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: 1 Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, 46, 544-560. [https://doi.org/10.1044/1092-4388\(2003/044\)](https://doi.org/10.1044/1092-4388(2003/044))
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19(4), 450-466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Daskalaki, E., Blom, E., Chondrogianni, V., & Paradis, J. (2020). Effects of parental input quality in child heritage language acquisition. *Journal of Child Language*, 47(4), 709-736. <https://doi.org/10.1017/S0305000919000850>
- Daskalaki, E., Chondrogianni, V., Blom, E., Argyri, F., & Paradis, J. (2019). Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research*, 35(3), 421-445. <https://doi.org/10.1177/0267658318787>
- Daskalaki, E., Soto-Corominas, A., Barisé, A., Paradis, J., Chen, X., & Gottardo, A. (2023). The interplay between syntactic and morphological comprehension in heritage contexts: The case of relative clauses in heritage Syrian Arabic. *Applied Psycholinguistics*, 44(6), 1043-1068. <https://doi.org/10.1017/S0142716423000401>
- Daskalaki, E., Soto-Corominas, A., Xia, V.Y., & Paradis, J. (2024). The role of parental characteristics, home language use, and schooling in children's Mandarin heritage language development in Canada. *Frontiers in Language Sciences*, 3, 1435200, <https://doi.org/10.3389/flang.2024.1435200>
- Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, 24(1), 3-24. <https://doi.org/10.1017/s0305000996002917>
- Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, 92(1), 25-45. <https://doi.org/10.1016/j.jecp.2005.03.007>
- Davies, C., Andrés-Roqueta, C., & Norbury, C. F. (2016). Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology*, 144, 98-113. <https://doi.org/10.1016/j.jecp.2015.11.011>
- de Almeida, L., Ferré, S., Barthez, M.-A., & dos Santos, C. (2019). What do monolingual and bilingual children with and without SLI produce when phonology is too complex? *First Language*, 39(2), 158-176. <https://doi.org/10.1177/0142723718805665>
- de Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism*, 7(3-4), 331-358. <https://doi.org/10.1075/lab.15019.alm>
- de Almeida, L., Rose, Y., & Freitas, J. (2012). Prosodic influence in bilingual phonological development: Evidence from a Portuguese French first language learner. In Biller A., Chung E. & Kimball A. (Eds.), *Proceedings of BUCLD 36* (pp. 42-52). Somerville: Cascadia Press.
- De Avila, E. A., & Duncan, S. E. (1980). The language minority child: A psychological, linguistic, and social analysis. In J. Alatis (Ed.), *Current issues in bilingual education* (pp. 104-137). Washington, DC: Georgetown University Press.
- de Bruin, A., Bak, T.H., & Della Sala, S. (2015a). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, 85, 15-26. <https://doi.org/10.1016/j.jml.2015.07.001>
- de Bruin, A., Treccani, B., & Della Sala, S. (2015b). Cognitive advantage in bilingualism: an example of publication bias?. *Psychological Science*, 26(1), 99-107. <https://doi.org/10.1177/0956797614557866>
- De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40(1), 119-151. <https://doi.org/10.1017/S0272263116000486>

- Declerck, M., & Philipp, A. M. (2015). A sentence to remember: Instructed language switching in sentence production. *Cognition*, 137, 166-173. <https://doi.org/10.1016/j.cognition.2015.01.006>
- Deevy, P., & Leonard, L. B. (2004). The Comprehension of Wh-Questions in Children With Specific Language Impairment. *Journal of Speech, Language & Hearing Research*, 47(4), 802-815. [http://doi.org/10.1044/1092-4388\(2004/060\)](http://doi.org/10.1044/1092-4388(2004/060))
- Degani, T., Kreiser, V., & Novogrodsky, R. (2019). The joint effects of bilingualism, DLD and item frequency on children's lexical-retrieval performance. *International Journal of Language & Communication Disorders*, 54(1), 1-15. <https://doi.org/10.1080/00207179.2018.1518181>
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Bristol: Multilingual Matters.
- De Houwer, A. (2014). The absolute frequency of maternal input to bilingual and monolingual children: A first comparison. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development*. (pp. 37-58). Amsterdam: John Benjamins.
- De Houwer, A. (2017). Bilingual language input environments, intake, maturity and practice. *Bilingualism: Language and Cognition*, 20(1), 19-20. <https://doi.org/10.1017/S1366728916000298>
- De Houwer, A. (2021). *Bilingual Development in Childhood*. Cambridge: Cambridge University Press.
- De Houwer, A., & Bornstein, M. H. (2016a). Balance patterns in early bilingual acquisition: a longitudinal study of word comprehension and production. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 134-155). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107375345.007>
- De Houwer, A., & Bornstein, M.H. (2016b). Bilingual Mothers' Language Choice in Child-directed Speech: Continuity and Change. *Journal of multilingual and multicultural development*, 37(7), 680-693. <https://doi.org/10.1080/01434632.2015.1127929>
- De Jong, J. (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure*. Doctoral Dissertation, Groningen University. Groningen: GRODIL.
- De Jong, J. (2004). Grammatical impairment: An overview and sketch of Dutch. In L. Verhoeven & H. v. Balkom (Eds.), *Classification of developmental language disorders: Theoretical issues and clinical implications* (pp. 261-281). Mahwah, NJ: Erlbaum.
- De Jong, J. (2008). Bilingualism and Language Impairment. In M.J. Ball, M.R. Perkins, N. Müller & S. Howard (Eds.), *The handbook in clinical linguistics* (pp. 261-274). Blackwell Publishing.
- De Jong, J. (2015). Elicitation task for subject-verb agreement. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing multilingual children. Disentangling bilingualism from language impairment* (pp. 25-37). Bristol: Multilingual Matters Ltd.
- Jong, J., Çavus, N., & Baker, A. (2010). Language Impairment in Turkish-Dutch Bilingual Children. In S. Topbaş & M. Yavaş (Ed.), *Communication Disorders in Turkish* (pp. 288-300). Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781847692474-018>
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, 22(4), 499-533.
- DeKeyser, R. (2012). Age effects in second language learning. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 442-460). Routledge: New York.
- Delage, H. (2008). *Evolution de l'hétérogénéité linguistique chez les enfants sourds moyens et légers: Etude de la complexité morphosyntaxique* [Doctoral Dissertation, François-Rabelais University, Tours].
- Delage, H., Durrleman, S., & Frauenfelder, U. H. (2016). Disentangling sources of difficulty associated with the acquisition of accusative clitics in French. *Lingua*, 180, 1-24. <https://doi.org/10.1016/j.lingua.2016.03.005>
- Delage, H., & Frauenfelder, U. H. (2019). Syntax and working memory in typically-developing children: Focus on syntactic complexity. *Language, Interaction and Acquisition*, 10(2), 141-176. <https://doi.org/10.1075/lia.18013.del>
- Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with Developmental Language Disorder. *Journal of Child Language*, 47(3), 600-632. <https://doi.org/10.1017/S0305000919000722>

- Delage, H., Monjauze, C., Hamann, C., & Tuller, L. (2008). Relative Clauses in Atypical Acquisition of French. In A. Gavarró & M. J. Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007* (pp. 166-176). Newcastle upon-Tyne: Cambridge Scholars Publishing.
- Delage, H., & Tuller, L. (2010). Evolution of syntactic complexity and avoidance strategies in children and adolescents with mild-to-moderate hearing loss. In A. Castro, M. Lobo, J. Costa, & F. Pratas, *Language acquisition and development: Proceedings of GALA 2009* (pp. 107-120). Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders*, 46(6), 613-627. <https://doi.org/10.1111/j.1460-6984.2011.00049.x>
- Delcenserie, A., & Genesee, F. (2017). The effects of age of acquisition on verbal memory in bilinguals. *International Journal of Bilingualism*, 21(5), 600-616. <https://doi.org/10.1177/1367006916639>
- Delcenserie, A., Genesee, F., & Gauthier, K. (2013). Language abilities of internationally adopted children from China during the early school years: Evidence for early age effects? *Applied Psycholinguistics*, 34(3), 541-568. <https://doi.org/10.1017/S0142716411000865>
- Delis, D.C., Kaplan, E., Kramer, J.H., & Ober, B.A. (2001). *Delis-Kaplan Executive Function Scale*. San Antonio, TX: Harcourt Assessment.
- de Lopéz, K. J., & Baker, A. E. (2015). Executive Functions in the Assessment of Bilingual Children with Language Impairment. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 277-298). Bristol: Multilingual Matters.
- Demir-Lira, Ö. E., Applebaum, L. R., Goldin-Meadow, S., & Levine, S. C. (2019). Parents' early book reading to children: Relation to children's later language and literacy outcomes controlling for other parent language input. *Developmental science*, 22(3), e12764. <https://doi.org/10.1111/desc.12764>
- Deml, I., Binder, K., Schulte, M., Merkert, A., Bien-Miller, L., Wildemann, A., & Schilcher, A. (2018). Eva-Prim. Evaluation im Primarbereich: Sprachförderung in alltäglichen und fachlichen Kontexten im Rahmen der Bund-Länder-Initiative BiSS. In S. Henschel, S. Gentrup, L. Beck & P. Stanat (Eds.), *Projektatlas evaluation: Erste ergebnisse aus den BiSS-Evaluationsprojekten* (pp. 32-34). Berlin: BiSS-Trägerkonsortium.
- Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, 12(1), 45-75. [https://doi.org/10.1016/0273-2297\(92\)90003-K](https://doi.org/10.1016/0273-2297(92)90003-K)
- Demuth, K. (2014). Prosodic licensing and the development of phonological and morphological representations. In A. Farris-Trimble & J. Barlow (Eds.), *Perspectives on phonological theory and acquisition: Papers in Honor of Daniel A. Dinnsen (Language Acquisition and Language Disorders Series, vol. 56, pp. 11-24)*. Amsterdam, The Netherlands: John Benjamins.
- Demuth, K. (2019). Prosodic constraints on children's use of grammatical morphemes. *First Language*, 39(1), 80-95. <https://doi.org/10.1177/0142723717751>
- Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 35(1), 99-127. <https://doi.org/10.1017/s0305000907008276>
- de Ribaupierre, A., & Bailleux, C. (1994). Developmental change in a spatial task of attentional capacity: An essay toward an integration of two working memory models. *International Journal of Behavioral Development*, 17(1), 5-35. <https://doi.org/10.1177/016502549401700102>
- Deuchar, M., & Muntz, R. (2003). Factors accounting for code-mixing in an early developing bilingual. In N. Müller (Ed.), *(In)vulnerable Domains in Multilingualism* (pp.161-190). Amsterdam: John Benjamins.
- de Villiers, J. (2003). Defining SLI: A linguistic perspective. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment* (pp. 425-447). Lawrence Erlbaum Associates.
- de Villiers, J., de Villiers, P.A., & Hoban, E. (1994). The central problem of functional categories in the

- English syntax of oral deaf children. In H. Tager-Flusberg (Ed.), *Constraints on Language Acquisition: Studies of Atypical Children* (pp. 9-47). Hillsdale, NJ.: Erlbaum.
- de Villiers, J., Kotfila, J., Roeper, T. (2020). When Is Recursion Easier for Children?. In Guijarro-Fuentes, P., Suárez-Gómez, C. (Eds.), *New Trends in Language Acquisition Within the Generative Perspective. Studies in Theoretical Psycholinguistics*, vol. 49 (pp. 239-256). Dordrecht: Springer. https://doi.org/10.1007/978-94-024-1932-0_10
- De Vincenzi, D., Arduino, L., Ciccarelli, L.S., & Job, R. (1999). Parsing strategies in children: comprehension of interrogative sentences. In S. Bagnara (Ed.), *Proceedings of the European Conference on Cognitive Science* (pp. 301-308). Rome: Istituto di Psicologia del CNR.
- Diamond A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature human behaviour*, 3(7), 692-701. <https://doi.org/10.1038/s41562-019-0609-3>
- Diessel, H. (2004). *The acquisition of complex sentences* (Vol. 105). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486531>
- Diessel, H. (2009). On the role of frequency and similarity in the acquisition of subject and non-subject relative clauses. In T. Givón & M. Shibatani (Eds.), *Syntactic Complexity: Diachrony, acquisition, neuro-cognition, evolution* (pp. 251-278). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.85.11ont>
- Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, 11(1-2), 131-151. <https://doi.org/10.1515/cogl.2001.006>
- Diessel, H., & Tomasello, M. (2005). Particle placement in early child language: A multifactorial analysis. *Corpus Linguistics and Linguistic Theory*, 1(1), 89-112. <https://doi.org/10.1515/cllt.2005.1.1.89>
- Dispaldro, M., Deevy, P., Altoé, G., Benelli, B., & Leonard, L. B. (2011). A cross-linguistic study of real-word and non-word repetition as predictors of grammatical competence in children with typical language development. *International Journal of Language & Communication Disorders*, 46(5), 564-578. <https://doi.org/10.1111/j.1460-6984.2011.00008.x>
- Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: a study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, 56(1), 323-336. [https://doi.org/10.1044/1092-4388\(2012/11-0304\)](https://doi.org/10.1044/1092-4388(2012/11-0304))
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). Young German children's early syntactic competence: a preferential looking study. *Developmental Science*, 11(4), 575-582. <https://doi.org/10.1111/j.1467-7687.2008.00703.x>
- Dittrich, K., Bossert, M.-L., Rothe-Wulf, A., & Klauer, K. C. (2017). The Joint Flanker Effect and the Joint Simon Effect: On the Comparability of Processes Underlying Joint Compatibility Effects. *Quarterly Journal of Experimental Psychology*, 70(9), 1808-1823. <https://doi.org/10.1080/17470218.2016.1207690>
- Dockrell, J. E., Messer, D., George, R., & Wilson, G. (1998). Children with word-finding difficulties--prevalence, presentation and naming problems. *International Journal of Language & Communication Disorders*, 33(4), 445-454. <https://doi.org/10.1080/136828298247721>
- Dockrell, J., Lindsay, G., Roulstone, S., & Law, J. (2014). Supporting children with speech, language and communication needs: an overview of the results of the Better Communication Research Programme. *International Journal of Language and Communication Disorders*, 49(5), 543-557. <https://doi.org/10.1111/1460-6984.12089>
- Dodd, B. (1995): *Differential diagnosis and treatment of children with speech disorder*. London: Whurr Publishers.
- Dodd, B. (2005): *Differential diagnosis and treatment of children with speech disorder*. London: Whurr Publishers.

- Dollaghan, C. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore: Paul H. Brookes.
- Dollaghan, C., Biber, M., & Campbell, T. (1995). Lexical influences on nonword repetition. *Lexical influences on nonword repetition. Applied Psycholinguistics*, 16(2), 211-222. <https://doi.org/10.1017/S0142716400007098>
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136-1146. <https://doi.org/10.1044/jslhr.4105.1136>
- Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: a meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, 54(4), 1077-1088. [https://doi.org/10.1044/1092-4388\(2010/10-0093\)](https://doi.org/10.1044/1092-4388(2010/10-0093))
- Domínguez, L., Arche, M. J., & Myles, F. (2011). Testing the predictions of the feature-assembly hypothesis: evidence from the L2 acquisition of Spanish aspect morphology. In N. Danis, K. Mesh & H. Sung (Eds.), *Proceedings of BUCLD 35* (pp. 183-196). Somerville: Cascadilla Press.
- Domínguez, L., Hicks, G., & Slabakova, R. (2019). Terminology choice in generative acquisition research: The case of “incomplete acquisition” in heritage language grammars. *Studies in Second Language Acquisition*, 41(2), 241-255. **Fehler! Linkreferenz ungültig.**
- Donders F. C. (1969). On the speed of mental processes. *Acta psychologica*, 30, 412-431. [https://doi.org/10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Döpke S. (1998). Competing language structures: the acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, 25(3), 555-584. <https://doi.org/10.1017/s0305000998003584>
- Döpke, S. (2000). *One Parent-One Language: An Interactional Approach*. Amsterdam: Benjamins.
- Dormann, C., Elith, J., Bacher, S., Bucmann, C., Carl, G., Carré, G.,::: S. Lautenbach. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance: open access. *Ecography*, 36(1), 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children’s phonology. *Language Acquisition: A Journal of Developmental Linguistics*, 25(1), 58-71. <https://doi.org/10.1080/10489223.2016.1243692>
- Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, 69(Suppl 1), 101-126. <https://doi.org/10.1111/lang.12322>
- Dromi, E., & Berman, R. A. (1986). Language-specific and language-general in developing syntax. *Journal of Child Language*, 13(2), 371-387. <https://doi.org/10.1017/s0305000900008114>
- Dromi, E., Leonard, L. B., & Shteyman, M. (1993). The grammatical morphology of Hebrew-speaking children with specific language impairment: some competing hypotheses. *Journal of Speech and Hearing Research*, 36(4), 760-771. <https://doi.org/10.1044/jslhr.3604.760>
- Dubois, P., St-Pierre, M. C., Desmarais, C., & Guay, F. (2020). Young Adults with Developmental Language Disorder: A Systematic Review of Education, Employment, and Independent Living Outcomes. *Journal of Speech, Language, and Hearing Research*, 63(11), 3786-3800. https://doi.org/10.1044/2020_JSLHR-20-00127
- Dulay, K. M., Tong, X., & McBride, C. (2017). The role of foreign domestic helpers in Hong Kong Chinese children’s English and Chinese skills: A longitudinal study. *Language Learning*, 67(2), 321-347. <https://doi.org/10.1111/lang.12222>
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: myth or reality?. *Experimental Psychology*, 61(3), 234-251. <https://doi.org/10.1027/1618-3169/a000243>
- Dunn, G. (2014). Statistics in psychiatry. In *International Encyclopedia of Statistical Science* (pp. 1136-1138). London: Arnold.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test (3rd ed.)*. Bloomington: Pearson Assessments.

- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (4th ed.)*. APA PsycTests. <https://doi.org/10.1037/t15144-000>
- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale (2nd ed.)*. Windsor, England: NFER-Nelson.
- Durkin, K., & Conti-Ramsden, G. (2007). Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. *Child Development*, 78(5), 1441-1457. <https://doi.org/10.1111/j.1467-8624.2007.01076.x>
- Durkin, K., & Conti-Ramsden, G. (2010). Young people with specific language impairment: A review of social and emotional functioning in adolescence. *Child Language Teaching and Therapy*, 26(2), 105-121. <https://doi.org/10.1177/0265659010368750>
- Durkin, K., Fraser, J., & Conti-Ramsden, G. (2010). School-age prework experiences of young people with a history of specific language impairment. *Journal of Special Education*, 45(2), 242-255. <https://doi.org/10.1177/0022466910362>
- Durkin, K., Conti-Ramsden, G., & Walker, A. J. (2011). Txt lang: Texting, textism use and literacy abilities in adolescents with and without specific language impairment. *Journal of Computer Assisted Learning*, 27(1), 49-57. <https://doi.org/10.1111/j.1365-2729.2010.00397.x>
- Durkin, K., Toseeb, U., Botting, N., Pickles, A., & Conti-Ramsden, G. (2017). Social Confidence in Early Adulthood Among Young People With and Without a History of Language Impairment. *Journal of Speech, Language, and Hearing Research*, 60(6), 1635-1647. https://doi.org/10.1044/2017_JSLHR-L-16-0256
- Duursma, E., Augustyn, M., & Zuckerman, B. (2008). Reading aloud to children: the evidence. *Archives of disease in childhood*, 93(7), 554-557. <https://doi.org/10.1136/adc.2006.106336>
- Dyson, A.T., & Amayreh, M.M. (2000). Phonological errors and sound changes in Arabic-speaking children. *Clinical Linguistics & Phonetics*, 14, 79-109. <https://doi.org/10.1080/026992000298850>
- Eadie, P., Nguyen, C., Carlin, J., Bavin, E., Bretherton, L., & Reilly, S. (2014). Stability of language performance at 4 and 5 years: measurement and participant variability. *International Journal of Language & Communication Disorders*, 49(2), 215-227. <https://doi.org/10.1111/1460-6984.12065>
- Ebbels, S., & van der Lely, H. (2001). Meta-syntactic therapy using visual coding for children with severe persistent SLI. *International Journal of Language & Communication Disorders*, 36 Suppl, 345-350. <https://doi.org/10.3109/13682820109177909>
- Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*. [https://doi.org/10.1044/1092-4388\(2011/10-0231\)](https://doi.org/10.1044/1092-4388(2011/10-0231))
- Ebert, K. D., Kohnert, K., Pham, G., Disher, J. R., & Payesteh, B. (2014). Three treatments for bilingual children with primary language impairment: examining cross-linguistic and cross-domain effects. *Journal of Speech, Language, and Hearing Research*, 57(1), 172-186. [https://doi.org/10.1044/1092-4388\(2013/12-0388\)](https://doi.org/10.1044/1092-4388(2013/12-0388))
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421-436. [https://doi.org/10.1044/1092-4388\(2004/034\)](https://doi.org/10.1044/1092-4388(2004/034))
- Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics*, 19(2), 279-309. <https://doi.org/10.1017/S0142716400010079>
- Eichler, N., Jansen, V., & Müller, N. (2013). Gender acquisition in bilingual children: French-German, Italian-German, Spanish-German and Italian-French. *International Journal of Bilingualism*, 17(5), 550-572. <https://doi.org/10.1177/1367006911435719>
- Eisenbeiss, S., Bartke, S., & Clahsen, H. (2006). Structural and Lexical Case in Child German: Evidence from Language-Impaired and Typically Developing Children. *Language Acquisition*, 13(1), 3-32. https://doi.org/10.1207/s15327817la1301_2

- Eisenbeiss, S., Narasimhan, B., & Voeikova, M. (2009). The Acquisition of Case. In: A. Malchukov und A. Spencer (Hg.): *The Oxford Handbook of Case*. Oxford: Oxford University Press, 369-383.
- Eisenberg, P. (1999). *Der Satz. Grundriß der deutschen Grammatik*. Stuttgart: J. B. Metzler.
- Eisenberg, P. (2013). *Grundriss der deutschen Grammatik* (vol. 4.). Stuttgart: J. B. Metzler.
- Eisenclas, S., Schalley, A., & Guillemin, D. (2013). The importance of literacy in the home language: the view from Australia. *Sage Open*, 3(4). <https://doi.org/10.1177/2158244013507270>
- Elbro, C., Borström, I., & Petersen, D.K. (1998). Predicting Dyslexia from Kindergarten: The Importance of Distinctness of Phonological Representations of Lexical Items. *Reading Research Quarterly*, 33, 36-60. <https://doi.org/10.1598/RRQ.33.1.3>
- Elbro, C., Dalby, M., & Maarbjerg, S. (2011). Language-learning impairments: a 30-year follow-up of language-impaired children with and without psychiatric, neurological and cognitive difficulties. *International Journal of Language & Communication Disorders*, 46(4), 437-448. <https://doi.org/10.1111/j.1460-6984.2011.00004.x>
- Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42(5), 1249-1260. <https://doi.org/10.1044/jslhr.4205.1249>
- Endo, Y., & Haegeman, L. (2019). Adverbial clauses and adverbial concord. *Glossa: a journal of general linguistics*, 4(1), 48. 1-32. <https://doi.org/10.5334/gjgl.589>
- Engel, P. M., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence?. *Journal of Speech, Language, and Hearing Research*, 51(6), 1580-1587. [https://doi.org/10.1044/1092-4388\(2008/07-0210\)](https://doi.org/10.1044/1092-4388(2008/07-0210))
- Engel de Abreu P. M. (2011). Working memory in multilingual children: is there a bilingual effect?. *Memory*, 19(5), 529-537. <https://doi.org/10.1080/09658211.2011.590504>
- Engel de Abreu, P. M., Baldassi, M., Puglisi, M. L., & Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, 56(2), 630-642. [https://doi.org/10.1044/1092-4388\(2012/12-0079\)](https://doi.org/10.1044/1092-4388(2012/12-0079))
- Engel de Abreu, P. M., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, 38, 552-561. <https://doi.org/10.1016/j.intell.2010.07.003>
- Engel de Abreu, P. M., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: enhanced cognitive control in low-income minority children. *Psychological Science*, 23(11), 1364-1371. <https://doi.org/10.1177/0956797612443836>
- Engel de Abreu, P. M., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, 104(4), 974-986. <https://doi.org/10.1037/a0028390>
- Engel de Abreu, P. M., Gathercole, S. E., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences*, 21(5), 569-574. <https://doi.org/10.1016/j.lindif.2011.06.002>
- Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1), 19-23. <http://www.jstor.org/stable/20182756>
- Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives on Psychological Science*, 13(2), 190-193. <https://doi.org/10.1177/1745691617720478>
- Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory*, vol. 44, (pp. 145-199). Elsevier Science.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309-331. <https://doi.org/10.1037/0096-3445.128.3.309>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in

- a nonsearch task. *Perception & Psychophysics*, 16(1), 143-149. <https://doi.org/10.3758/BF03203267>
- Evans, J. L., Gillam, R. B., & Montgomery, J. W. (2018). Cognitive predictors of spoken word recognition in children with and without developmental language disorders. *Journal of Speech, Language, and Hearing Research*, 61(6), 1409-1425.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 321-335. [https://doi.org/10.1044/1092-4388\(2009/07-0189\)](https://doi.org/10.1044/1092-4388(2009/07-0189))
- Everitt, A., Hannaford, P., & Conti-Ramsden, G. (2013). Markers for persistent specific expressive language delay in 3-4-year-olds. *International Journal of Language & Communication Disorders*, 48(5), 534-553. <https://doi.org/10.1111/1460-6984.12028>
- Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics & Phonetics*, 26(2), 148-163. <https://doi.org/10.3109/02699206.2011.595526>
- Fabiano-Smith, L., & Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, 53(1), 160-178. [https://doi.org/10.1044/1092-4388\(2009/07-0064\)](https://doi.org/10.1044/1092-4388(2009/07-0064))
- Fahim, D. (2005). *Developmental language impairment in Egyptian Arabic* [Unpublished thesis, Birkbeck, University of London].
- Fahim, D. (2017). Verb morphology in Egyptian Arabic developmental language impairment. *Arab Journal of Applied Linguistics*, 2(1), 49-73.
- Falcaro, M., Pickles, A., Newbury, D. F., Addis, L., Banfield, E., Fisher, S. E., Monaco, A. P., Simkin, Z., Conti-Ramsden, G., & SLI Consortium (2008). Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment. *Genes, brain, and behavior*, 7(4), 393-402. <https://doi.org/10.1111/j.1601-183X.2007.00364.x>
- Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics*, 32(4), 711-738. <https://doi.org/10.1017/S0142716411000038>
- Farrant, B. M., Maybery, M. T., & Fletcher, J. (2012). Language, cognitive flexibility, and explicit false belief understanding: Longitudinal analysis in typical development and specific language impairment. *Child Development*, 83(1), 223-235.
- Fassbender, I., & Leyendecker, B. (2018). Socio-economic status and psychological well-being in a sample of Turkish immigrant mothers in Germany. *Frontiers in psychology*, 9, 1586. <https://doi.org/10.3389/fpsyg.2018.01586>
- Fassi Fehri, A. (2004). Tense/aspect interaction and variation. In J. Guéron & J. Lecarme (Eds.), *The syntax of time* (pp. 235-257). Cambridge: MIT Press.
- Fatzer, S. T., & Roebers, C. M. (2012). Language and executive functions: The effect of articulatory suppression on executive functioning in children. *Journal of Cognition and Development*, 13(4), 454-472.
- Faust, M., Dimitrovsky, L., & Davidi, S. (1997). Naming difficulties in language-disabled children: preliminary findings with the application of the tip-of-the-tongue paradigm. *Journal of Speech, Language, and Hearing Research*, 40(5), 1026-1036. <https://doi.org/10.1044/jslhr.4005.1026>
- Ferguson, C. (1959). Diglossia. *WORD*, 15(2), 325-340. <https://doi.org/10.1080/00437956.1959.11659702>
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234-248.
- Ferré, S., dos Santos, C., & de Almeida, L. (2015). Potential Phonological Markers for SLI in Bilingual Children. In E. Grillo & K. Jenson (Eds.), *Proceedings of BUCLD 39* (pp. 152-164). Somerville, MA: Cascadilla Press.
- Ferré, S., Tuller, L., Sizaret, E., & Barthez, M.-A. (2012). Acquiring and avoiding phonological complexity in SLI vs. typical development of French: the case of consonant clusters. In P. Hoole, L. Bombine, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant Clusters and Structural Complexity*. Berlin: de Gruyter (pp. 285-308). <https://doi.org/10.1515/9781614510772.285>

- Féry, C. (2008). *Phonologie des Deutschen. Eine optimalitätstheoretische Einführung*. Potsdam: University of Potsdam.
- Feyter, J. J. de, & Winsler, A. (2009). The early developmental competencies and school readiness of low-income, immigrant children: Influences of generation, race/ethnicity, and national origins. *Early Childhood Research Quarterly*, 24(4), 411-431.
- Fichman, S., Walters, J., Melamed, R., & Altman, C. (2022). Reference to characters in narratives of Russian-Hebrew bilingual and Russian and Hebrew monolingual children with developmental language disorder and typical language development. *First Language*, 42(2), 263-291. <https://doi.org/10.1177/0142723720962938>
- Field, A. (2018). *Discovering statistics using SPSS*. (5th Ed.). Sage Publications.
- Fikkert, J. P. M., & Freitas, M. J. (2004). The role of language-specific phonotactics in the acquisition of onset clusters. In L. Cornips & J. Doetjes (Eds.), *Linguistics in the Netherlands 2004* (pp. 58-68). Amsterdam, John Benjamins.
- Fikkert, P., & Penner, Z. (1998). Stagnation in prosodic development of language disordered children. In A. Greenhill (Ed.), *The proceedings of the 22nd Boston University conference on language development* (pp. 201-212). Somerville: Cascadilla Press.
- Finneran, D., Francis, A., & Leonard, L. (2009). Sustained attention in children with specific language impairment. *Journal of Speech and Hearing Research*, 52, 915-929.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80(1), 27-38. <https://doi.org/10.1093/biomet/80.1.27>
- Fisher, S. E. (2005). Dissection of molecular mechanisms underlying speech and language disorders. *Applied Psycholinguistics*, 26(1), 111-128.
- Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, 18(2), 168-170. <https://doi.org/10.1038/ng0298-168>
- Flax, J.F., Realpe-Bonilla, T., Hirsch, L.S., Brzustowicz, L.M., Barlett, C.W., & Tallal, P. (2003). Specific Language Impairment in families: Evidence from co-occurrence with reading impairments. *Journal of Speech, Language, and Hearing Research*, 46, 530-543.
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., & Zebib, R. (2018). How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, 25(1), 85-101. <https://doi.org/10.1080/10489223.2016.1192635>
- Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, 34, 153-175.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125-3134.
- Flores, C. (2015). Losing a language in childhood: a longitudinal case study on language attrition. *Journal of Child Language*, 42(3), 562-590. <https://doi.org/10.1017/S0305000914000233>
- Flores, C. (2017). Problematizing the scope of language attrition from the perspective of bilingual returnees. *Linguistic Approaches to Bilingualism*, 7(6), 691-695. <https://doi.org/10.1075/lab.00006.flo>
- Flores, C., & Rinke, E. (2020). The relevance of language-internal variation in predicting heritage language grammars. *Bilingualism: Language and Cognition*, 23(1), 25-26. <https://doi.org/10.1017/S1366728919000464>
- Flores, C., Santos, A. L., Jesus, A., & Marques, R. (2017). Age and input effects in the acquisition of mood in Heritage Portuguese. *Journal of Child Language*, 44(4), 795-828. <https://doi.org/10.1017/S0305000916000222>
- Fodor, F. J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press
- Foley, C., & Flynn, S. (2013). The role of the native language. In J. Herschensohn, & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 97-113). Cambridge, UK: Cambridge University Press.

- Forrest, C. L., Gibson, J. L., Halligan, S. L., & St Clair, M. C. (2018). A longitudinal analysis of early language difficulty and peer problems on later emotional difficulties in adolescence: evidence from the Millennium Cohort Study. *Autism & Developmental Language Impairments*, 3. <https://doi.org/10.1177/2396941518795392>
- Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 97-118). Hillsdale, NJ: Routledge.
- Fox, A., & Dodd, B. (1999). Der Erwerb des phonologischen Systems in der deutschen Sprache. *Sprache Stimme Gehör*, 23, 183-191.
- Fox, B. A., & Thompson, S. A. (1990). A Discourse Explanation of the Grammar of Relative Clauses in English Conversation. *Language*, 66(2), 297-316. <https://doi.org/10.2307/414888>
- Fox, D., & Grodzinsky, Y. (1998). Children's passive: A view from the by-phrase. *Linguistic Inquiry*, 29, 311-332. <https://doi.org/10.1162/002438998553761>
- Fox-Boyer, A. V. (2014). *Psycholinguistischer Analyse kindlicher Aussprachestörungen-II: PLAKSS II*. Frankfurt: Pearson.
- Fox-Boyer, A. (2016). *Kindliche Aussprachestörungen: Phonologischer Erwerb, Differenzialdiagnostik, Therapie*. Idstein: Schulz-Kirchner.
- Fox-Boyer, A., Fricke, S., & Albrecht, K. M. (2020). Phonologische Prozesse in der deutschen Sprache türkisch-deutsch bilingualer Kinder. *Sprache-Stimme Gehör*, 44(04), 216-223. DOI: 10.1055/a-1127-5908
- Fox-Boyer, A., & Salgert, K. (2014). Erwerb und Störungen der Aussprache bei mehrsprachigen Kindern. In S. Chilla & S. Haberzettl (Eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen – Mehrsprachigkeit* (pp. 109–121). München: Elsevier.
- Franck, J., & Delage, H. (2022). The Interplay of Emotions, Executive Functions, Memory and Language: Challenges for Refugee Children. *Languages*, 7(4), 309. <https://doi.org/10.3390/languages7040309>
- Frazier, L., & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331-344. [http://doi.org/10.1016/0749-596X\(89\)90037-5](http://doi.org/10.1016/0749-596X(89)90037-5)
- Fredman, M. (2006). Recommendations for working with bilingual children—Prepared by the multilingual affairs committee of IALP. *Folia Phoniatrica et Logopaedica*, 58, 458-464.
- Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2017). Cross-linguistic phonotactic competition and cognitive control in bilinguals. *Journal of Cognitive Psychology*, 29(7), 783-794. <https://doi.org/10.1080/20445911.2017.1321553>
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17(2), 172-179. <https://doi.org/10.1111/j.1467-9280.2006.01681.x>
- Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1), 67-88. <https://doi.org/10.1016/j.lingua.2008.09.002>
- Friedmann, N., & Costa, J. (2010). The child heard a coordinated sentence and wondered: on children's difficulty in understanding coordination and relative clauses with crossing dependencies. *Lingua*, 120, 1502-1515.
- Friedmann, N., & Haddad-Hanna, M. (2014). Types of developmental dyslexia in Arabic. In E. Saiegh-Haddad & M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives. Language and Literacy Series* (pp. 119-152). Amsterdam: Springer.
- Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development. *Journal of Child Language*, 31(3), 661-681. <https://doi.org/10.1017/s0305000904006269>
- Friedmann, N., & Novogrodsky, R. (2008). Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI. In A. Gavarró & M. J. Freitas (Eds.), *Language Acquisition and Development* (pp. 205-217). Cambridge: Cambridge University.

- Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI. *Lingua*, 121(3), 367-382. <https://doi.org/10.1016/j.lingua.2010.10.004>
- Friedmann, N., & Reznick, J. (2021). Stages rather than ages in the acquisition of movement structures: Data from sentence repetition and 27696 spontaneous clauses. *Glossa: a journal of general linguistics*, 6(1): 143, 1-69. <https://doi.org/10.16995/glossa.5716>
- Friedmann, N., Rizzi, L., & Belletti, A. (2017). No case for Case in locality: Case does not help interpretation when intervention blocks A-bar chains. *Glossa: a journal of general linguistics*, 2(1). <https://doi.org/10.5334/gjgl.165>
- Friedmann, N., & Szterman, R. (2011). The comprehension and production of Wh-questions in deaf and hard-of-hearing children. *Journal of Deaf Studies and Deaf Education*, 16(2), 212-235.
- Friedmann, N., Yachini, M., & Szterman, R. (2015). Relatively easy relatives: children with syntactic SLI avoid intervention. In E. Di Domenico, C. Hamann, & S. Matteini (Eds.), *Structures, Strategies and Beyond. Studies in Honour of Adriana Belletti* (pp. 303-320). Amsterdam: John Benjamins. <https://doi.org/10.1075/la.223.14fri>
- Frijters, J. C., Barron, R. W., & Brunello, M. (2000). Direct and mediated influences of home literacy and literacy interest on prereaders' oral vocabulary and early written language skill. *Journal of Educational Psychology*, 92(3), 466-477. <https://doi.org/10.1037/0022-0663.92.3.466>
- Frizelle, P., & Fletcher, P. (2015). The role of memory in processing relative clauses in children with specific language impairment. *American Journal of Speech-Language Pathology*, 24, 47-59.
- Froud, K., & van der Lely, H. K. (2008). The count-mass distinction in typically developing and grammatically specifically language impaired children: new evidence on the role of syntax and semantics. *Journal of Communication Disorders*, 41(3), 274-303. <https://doi.org/10.1016/j.jcomdis.2007.11.001>
- Gärtner, H. M. (2014). On Covert Modality in German Root Infinitives. In *Proceedings of the 31st West Coast Conference on Formal Linguistics* (pp. 199-206). Somerville: Cascadia Press.
- Gagarina, N., Grimm, A., Rinker, T., & Scherger, Anna-Lena (2026). LITMUS-Batterie als domänenübergreifendes Instrument zur Identifikation von Sprachentwicklungsstörungen bei mehrsprachigen Kindern in Deutschland. *Sprache Stimme Gehör*. DOI: 10.1055/a-2736-9425
- Gagarina, N., & Klassert, A. (2018). Input dominance and development of home language in Russian-German bilinguals. *Frontiers in Communication*, 3, 40. <https://doi.org/10.3389/fcomm.2018.00040>
- Gagarina, N., Klassert, A., and Topaj, N. (2010). Russian Language Proficiency Test for Multilingual Children. *ZAS Papers in Linguistics*, 54, Berlin: ZAS.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of Narrative Abilities in Bilingual Children. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 243-276). Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781783093137-011>
- Gagarina, N., Klop, D., Tsimpli, I., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, 37(1), 11-17. <https://doi.org/10.1017/S0142716415000399>
- Galal, M. (2004). Relativizer 'illi' in Arabic dialects. *Kansas Working Papers in Linguistics*, 27, 44-65.
- Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 57(4), 1363-1382.
- Gallon, N. (2002). *Prosodic complexity in specific language impairment* [Master's thesis, University College London], London, England.
- Gallon, N., H. J., & Van der Lely, H. (2007). Non-word repetition: an investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics & Phonetics*. 21(6), 445-455.
- Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes*, 13(2-3), 269-306. <https://doi.org/10.1080/016909698386537>

- Gathercole, S. E. (2006a). Complexities and constraints in nonword repetition and word learning. *Applied Psycholinguistics*, 27(4), 599-613. <https://doi.org/10.1017/S014271640606053X>
- Gathercole, S. E. (2006b). Nonword repetition and word learning: the nature of the relationship. *Applied Psycholinguistics*, 27, 513-543.
- Gathercole, S. E., & Adams, A. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, 33, 672-688.
- Gathercole, S. E., & Alloway, T.P. (2008). *Working memory and learning: A teacher's guide*. London: Sage Publications.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28(2), 200-213. [https://doi.org/10.1016/0749-596X\(89\)90044-2](https://doi.org/10.1016/0749-596X(89)90044-2)
- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *Journal of memory and language*, 29(3), 336-360. [https://doi.org/10.1016/0749-596X\(90\)90004-J](https://doi.org/10.1016/0749-596X(90)90004-J)
- Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, 8(3), 259-272. <https://doi.org/10.1007/BF03174081>
- Gathercole, S. E., Brown, L., & Pickering, S.J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational Psychology*, 70, 177-194.
- Gathercole, S. E., Frankish, C., Pickering, S., & Peaker, S. (1999). Phonotactic influences on short term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 84-95.
- Gathercole, S. E., Pickering, S., Ambridge, B., & Wearing, H. (2004a). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177-90.
- Gathercole, S., Pickering, S., Knight, C., & Stegmann, Z. (2004b). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18, 1-16. <https://doi.org/10.1002/acp.934>
- Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28(5), 887-898. <https://doi.org/10.1037/0012-1649.28.5.887>
- Gathercole, V. C. M. (2002a). Command of the mass/count distinction in bilingual and monolingual children: An English morphosyntactic distinction. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 175-206). Clevedon: Multilingual Matters.
- Gathercole, V. C. M. (2002b). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 207-219). Clevedon: Multilingual Matters.
- Gathercole, V. C. M. (2002c). Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 220-254). Clevedon: Multilingual Matters.
- Gathercole, V. C. M. (2007). Miami and North Wales, so far and yet so near: Constructivist account of morpho-syntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism*, 10, 224-247. <http://dx.doi.org/10.2167/beb442>
- Gathercole, V. C. M. (2014). Bilingualism matters: One size does not fit all. *International Journal of Behavioral Development*, 38(4), 359-366.
- Gathercole, V. C. M. (2016). Factors moderating proficiency in bilingual speakers. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 123-140). American Psychological Association. <https://doi.org/10.1037/14939-008>
- Gathercole, V. C. M., Kennedy, I., & Thomas, E. M. (2016). Socioeconomic level and bilinguals' performance on language and cognitive measures. *Bilingualism: Language and Cognition*, 19(5), 1057-1078. <https://doi.org/10.1017/S1366728915000504>

- Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12(2), 213-237. <https://doi.org/10.1017/S1366728909004015>
- Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas-Guasch, N., ... & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in psychology*, 5, 11. <https://doi.org/10.3389/fpsyg.2014.00011>
- Gathercole, V., Thomas, E., Roberts, E., Hughes, C., & Hughes, E. (2013). 2. Why Assessment Needs to Take Exposure into Account: Vocabulary and Grammatical Abilities in Bilingual Children. In V. Gathercole (Ed.), *Issues in the Assessment of Bilinguals* (pp. 20-55). Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781783090105-004>
- Gaulin, C. A., & Campbell, T. F. (1994). Procedure for Assessing Verbal Working Memory in Normal School-Age Children: Some Preliminary Data. *Perceptual and Motor Skills*, 79(1), 55-64. <https://doi.org/10.2466/pms.1994.79.1.55>
- Gavarró, A. (2003). Economy and word order patterns in bilingual English-Dutch acquisition. *Bilingualism: Language and Cognition*, 6(1), 69-79. <https://doi.org/10.1017/S1366728903001044>
- Gavarró, A., & Fortón, N. (2014). Person features and the acquisition of clitics. In C. Contemori & L. Dal Pozzo (Eds.), *Inquiries into Linguistic Theory and Language Acquisition. Papers offered to Adriana Belletti, Siena*. CISCL Press.
- Gavruseva, E., & Lardiere, D. (1996). The emergence of extended phrase structure in child L2 acquisition. In A. Stringfellow, D. Cahana-Amitay, E. Hughes & A. Zukowski (Eds.), *Proceedings of BUCLD 20* (pp. 225-236). Somerville, MA: Cascadilla Press.
- Gawlitsek-Maiwald, I., & Tracy, R. (1996). Bilingual bootstrapping. *Linguistics*, 34(5), 901-926. <https://doi.org/10.1515/ling.1996.34.5.901>
- Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized Phrase Structure Grammar*. Cambridge: Harvard University Press.
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, 88, 121-151. <https://doi.org/10.1016/j.jecp.2004.03.002>
- Genesee, F. (2015). Myths about early childhood bilingualism. *Canadian Psychology/Psychologie canadienne*, 56(1), 6-15. <https://doi.apa.org/doi/10.1037/a0038599>
- Genesee, F., & Nicoladis, E. (2007). Bilingual first language acquisition. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 324-342). Oxford: Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470757833.ch16>
- Genesee, F., Paradis, J., & Crago, M. (2004). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*. Baltimore, MD: Brookes Publishing.
- Genevska-Hanke, D., & Hamann, C. (2024). Overt and null subjects in child heritage Bulgarian: The role of language dominance. In H. A. Althagafi & J. Ray (Eds.), *Proceedings of BUCLD 48* (pp. 195-208). Somerville: Cascadilla Press.
- Ghawi-Dakwar, O., & Saiegh-Haddad, E. (2025). Word Learning in Arabic Diglossia in Children With Typical Language Development and Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, 68(3S), 1533-1551. https://doi.org/10.1044/2024_JSLHR-23-00618
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1), 1-76. [https://doi.org/10.1016/S0010-0277\(98\)00034-1](https://doi.org/10.1016/S0010-0277(98)00034-1)
- Gibson, T. A., Summers, C., Peña, E. D., Bedore, L. M., Gillam, R. B., & Bohman, T. M. (2015). The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism*, 18(3), 551-560. <https://doi.org/10.1017/s1366728914000248>
- Gillam, R. B., Montgomery, J. W., Evans, J. L., & Gillam, S. L. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: Implications for

- assessment and treatment. *International Journal of Speech-Language Pathology*, 21(3), 240-251. <https://doi.org/10.1080/17549507.2018.1559883>
- Gillon, G. (2002). Phonological awareness intervention for children: From the research laboratory to the clinic. *ASHA Leader*, 7(22), 4-17. <https://doi.org/10.1044/leader.FTR2.07222002>.
- Gillon, G. (2004). *Phonological awareness: From research to practice*. New York: The Guilford Press.
- Girbau, D. (2016). The Non-word Repetition Task as a clinical marker of Specific Language Impairment in Spanish-speaking children. *First Language*, 36(1), 30-49. <https://doi.org/10.1177/0142723715626069>
- Girbau, D., & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, 42(1), 59-75
- Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish-English bilingual children with and without specific language impairment. *Journal of Communication Disorders*, 41(2), 124-145. <https://doi.org/10.1016/j.jcomdis.2007.07.001>
- Gleitmann, L. (1990). The structural sources of verb meaning. *Language Acquisition*, 1(1), 3-55. https://doi.org/10.1207/s15327817la0101_2
- Glick, J. E., Walker, L., & Luz, L. (2013). Linguistic isolation in the home and community: protection or risk for young children? *Social Science Research*, 42(1), 140-154.
- Glück, C. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6-10* (2nd ed.) Munich: Elsevier, Urban & Fischer.
- Goad, H., & Rose, Y. (2004). Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp.109-157). Cambridge: Cambridge University Press.
- Goldberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*, 29(1), 41-65.
- Goldstein, B. A., Bunta, F., Lange, J., Rodriguez, J., & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. *American Journal of Speech-Language Pathology*, 19(3), 238-247. [https://doi.org/10.1044/1058-0360\(2010/08-0086\)](https://doi.org/10.1044/1058-0360(2010/08-0086))
- Gollan, T. H., Starr, J., & Ferreira, V. S. (2015). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review*, 22(1), 147-155. <https://doi.org/10.3758/s13423-014-0649-7>
- Goodluck, H. (2010). Object extraction is not subject to child relativized minimality. *Lingua*, 120(6), 1516-1521. <https://doi.org/10.1016/j.lingua.2009.10.005>
- Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature*, 344(6268), 715.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology. Learning, memory, and cognition*, 27(6), 1411-1423. <https://doi.org/10.1037/0278-7393.27.6.1411>
- Gou, Z., Choudhury, N., & Benasich, A. A. (2011). Resting frontal gamma power at 16, 24 and 36 months predicts individual differences in language and cognition at 4 and 5 years. *Behavioural Brain Research*, 220(2), 263-270. <https://doi.org/10.1016/j.bbr.2011.01.048>.
- Govindarajan, K., & Paradis, J. (2019). Narrative abilities of bilingual children with and without Developmental Language Disorder (SLI): Differentiation and the role of age and input factors. *Journal of Communication Disorders*, 77, 1-16.
- Graf Estes, K., Evans, J. L., & Else-Quest, N. (2007). Differences in nonword repetition performance for children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177-195. [https://doi.org/10.1044/1092-4388\(2007/015\)](https://doi.org/10.1044/1092-4388(2007/015))

- Granfeldt, J., Schlyter, S., & Kihlstedt, M. (2007). French as cL2, 2L1 and L1 in pre-school children. *PERLES: petites études romanes de Lund. Extra seriem: commentaires & communications*, 21, 6-43.
- Gray S. (2004). Word learning by preschoolers with specific language impairment: predictors and poor learners. *Journal of Speech, Language, and Hearing Research*, 47(5), 1117-1132. [https://doi.org/10.1044/1092-4388\(2004/083\)](https://doi.org/10.1044/1092-4388(2004/083))
- Gray, S., Pittman, A., & Weinhold, J. (2014). Effect of phonotactic probability and neighborhood density on word-learning configuration by preschoolers with typical development and specific language impairment. *Journal of Speech, Language, and Hearing Research*, 57(3), 1011-1025. https://doi.org/10.1044/2014_JSLHR-L-12-0282
- Gray, S., Reiser, M., & Brinkley, S. (2012). Effect of onset and rhyme primes in preschoolers with typical development and specific language impairment. *Journal of Speech, Language, and Hearing Research*, 55(1), 32-44. [https://doi.org/10.1044/1092-4388\(2011/10-0203\)](https://doi.org/10.1044/1092-4388(2011/10-0203))
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1, 67-81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25, 515-530.
- Grewendorf, G. (1988). *Aspekte der deutschen Syntax*. Tübingen: Narr.
- Grießhaber, W. (2013). Die Profilanalyse für Deutsch als Diagnoseinstrument zur Sprachförderung. *Stiftung Mercator: proDaZ*. Universität Essen.
- Grijzenhout, J., & Penke, M. (2005). On the interaction of phonology and morphology in language acquisition and German and Dutch Broca's Aphasia: the case of inflected verbs. In *Yearbook of morphology 2005* (pp. 49-81). Dordrecht: Springer.
- Grillo, N. (2008). *Generalized minimality: syntactic underspecification in Broca's aphasia* [Doctoral dissertation]. Utrecht: LOT Publications. <http://www.lotpublications.nl/generalized-minimality-generalized-minimality-syntactic-underspecification-in-brocas-aphasia>
- Grillo, N. (2009). Generalized minimality: feature impoverishment and comprehension deficits in agrammatism. *Lingua*, 119(10), 1426-1443. <https://doi.org/10.1016/j.lingua.2008.04.003>
- Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet*, 365(9469), 1500-1505. [https://doi.org/10.1016/S0140-6736\(05\)66422-7](https://doi.org/10.1016/S0140-6736(05)66422-7)
- Grimm, A. (2022). The Use of the LITMUS Quasi-Universal Nonword Repetition Task to Identify DLD in Monolingual and Early Second Language Learners Aged 8 to 10. *Languages*, 7(3), 218. <https://doi.org/10.3390/languages7030218>
- Grimm, A., & Domahs, U. (2025). The acquisition of consonant clusters and word stress by early second language learners of German: Evidence for cross-linguistic influence?. *Linguistic Approaches to Bilingualism*, 15(2), 212-242. <https://doi.org/10.1075/lab.21026.gri>
- Grimm, A., Ferré, S., dos Santos, C., & Chiat, S. (2014). Can nonwords be language-independent? Crosslinguistic evidence from monolingual and bilingual acquisition of French, German, and Lebanese. In *Symposium on Language Impairment Testing in Multilingual Settings (LITMUS): Disentangling bilingualism and SLI*, Amsterdam, July 13-19, 2014.
- Grimm, A., & Hübner, J. (forthcoming). Nonword repetition by bilingual learners of German: The role of language-specific complexity. In C. dos Santos & L. de Almeida, *Bilingualism and Specific Language Impairment, Bi-SLI 2015*. Amsterdam: Benjamins.
- Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research*, 7(4), 821-841. <https://doi.org/10.1007/s12187-013-9230-6>
- Grimm, A., & Schulz, P. (2016). Warum man bei mehrsprachigen Kindern dreimal nach dem Alter fragen sollte: Sprachfähigkeiten simultan-bilingualer Lerner im Vergleich mit monolingualen und frühen Zweitsprachlernern. *Diskurs Kindheits- und Jugendforschung*, 11(1), 27-42. <https://doi.org/10.3224/diskurs.v11i1.22247>
- Grimm, A., & Schulz, P. (2017). Die Rolle von Risikofaktoren für die Diagnose von SSES bei ein- und

- mehrsprachigen Kindern. *Spektrum Patholinguistik*, 10, 21-49
- Grimm, A., & Schulz, P. (2021). Phonology and sentential semantics: markers of SLI in bilingual children at age 6? In K. Grohmann & S. Armon-Lotem (Eds.), *LITMUS in action: Comparative studies across Europe, TILAR* (pp. 263-300). Amsterdam: Benjamin. <https://doi.org/10.1075/tilar.29>
- Grimm, H. (1986). Ontogenese der Sprache als Fortsetzung nichtsprachlichen Handelns. In H. G. Bosshardt (Ed.), *Perspektiven auf Sprache. Interdisziplinäre Beiträge zum Gedenken an Hans Hörmann* (pp. 166-184). Berlin: De Gruyter.
- Grimm, H. (1993). Syntax and Morphological Difficulties in German-Speaking Children with Specific Language Impairment: Implications for Diagnosis and Intervention. In H. Grimm & H. Skowronek (Eds.): *Language Acquisition Problems and Reading Disorders Aspects of Diagnosis and Intervention* (pp. 25-64). Berlin: De Gruyter.
- Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen [A Test of Language Development for Three to Five Year Old Children. Diagnosis of Language Processing Skills and Auditory Memory Performance]*. Göttingen: Hogrefe.
- Grimm, H. (2003). *Sprachscreening für das Vorschulalter (SSV) [A Language Screening for Preschool Children]*. Göttingen: Hogrefe.
- Grimm, H., & Weinert, S. (1990). Is the syntax development of dysphasic children deviant and why? New findings to an old question. *Journal of Speech and Hearing Research*, 33(2), 220-228. <https://doi.org/10.1044/jsr.3302.220>
- Grinstead, J., Baron, A., Vega-Mendoza, M., De la Mora, J., Cantú-Sánchez, M., & Flores, B. (2013). Tense marking and spontaneous speech measures in Spanish specific language impairment: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 56, 352-363.
- Grinstead, J., Lintz, P., Vega-Mendoza, M., De la Mora, J., Cantú-Sánchez, M., & Flores-Avalos, B. (2014). Evidence of optional infinitive verbs in the spontaneous speech of Spanish-speaking children with SLI. *Lingua*, 140, 52-66.
- Grohmann, K., & Armon-Lotem, S. (Eds.) (2021). *Language Impairment in Multilingual Settings: LITMUS in action across Europe, TILAR*. Amsterdam: John Benjamins Publishing Company.
- Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development*, 6, 467-477.
- Grosjean, F. (1992). Another view of bilingualism. In R. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 51-62). The Netherlands: Elsevier.
- Grosjean, F. (1997). The bilingual individual. *Interpreting*, 2(1-2), 163-187.
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/9780674056459>
- Grosjean, F. (2016). The complementarity principle and its impact on processing, acquisition and dominance. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues in Measurement and Operationalization* (pp. 66-84). Cambridge: Cambridge University Press.
- Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, 33, 325-40.
- Grüter, T. (2006). *Object clitics and null objects in the acquisition of French*. Unpublished doctoral dissertation, McGill University.
- Grüter, T., & Crago, M., (2012). Object clitics and their omission in child L2 French: the contributions of processing limitations and L1 transfer. *Bilingualism: Language and Cognition*, 15, 531-549.
- Grüter, T., & Paradis, J. (Eds.). (2014). *Input and experience in bilingual development*. Amsterdam, The Netherlands: John Benjamins.
- Guasti, M.-T. (1994). Verb syntax in Italian child grammar: Finite and non-finite verbs. *Language Acquisition*, 3, 1-40.
- Guasti, M. T. (2002). *Language Acquisition: The Growth of Grammar*. Cambridge, Mass.: MIT Press.

- Guasti, M. T., Branchini, C., & Arosio, F. (2012). Interference in the production of Italian subject and object wh-questions. *Applied Psycholinguistics*, 33(1), 185-223. <http://dx.doi.org/10.1017/S0142716411000324>
- Guasti, M. T., & Cardinaletti, A. (2003). Relative clause formation in Romance child's production. *Probus*, 15(1), 47-89. <https://doi.org/10.1515/prbs.2003.005>
- Guasti, M.-T., Stavrakaki, S., & Arosio, F. (2008). Number and case in the comprehension of relative clauses: evidence from Italian and Greek. In A. Gavarró & M.J. Freitas (Eds.), *Language Acquisition and Development* (pp. 230-240). Cambridge: Cambridge Scholars Press/CSP.
- Guiberson, M., & Atkins, J. (2012). Speech-language pathologists' preparation, practices, and perspectives on serving culturally and linguistically diverse children. *Communication Disorders Quarterly*, 33(3), 169-180. <https://doi.org/10.1177/1525740110384132>
- Gussmann, E. (2007). *The phonology of Polish*. Oxford University Press.
- Gutiérrez-Clellen, V. (2012). Narrative development and disorders in bilingual children. In B. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (2nd ed., pp. 233-250). Baltimore, MD: Brookes Publishing.
- Gutiérrez-Clellen, V., Simon-Cerejido, G., & Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics*, 29, 3-19.
- Gwinner, A., Gaglia, S., & Grijzenhout, J. (2012). The effect of prosody on the acquisition of morphemes: An experimental study with German, Italian and German-Italian children. In M. G. Busà & A. Stella (Eds.), *Methodological Perspectives on Second Language Prosody* (pp. 37-41). Papers from ML2P 2012, CLEUP, Padova.
- Haberzettl, S. (2003). 'Tinkering' with chunks: Form-oriented strategies and idiosyncratic utterance patterns without functional implications in the IL of Turkish speaking children learning German. In C. Dimroth & M. Starren (Eds.), *Information structure, linguistic structure and the dynamics of language acquisition* (pp. 45-63). John Benjamins. <https://doi.org/10.1075/sibil.26.04hab>
- Haberzettl, S. (2005). *Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache*. Tübingen: Max Niemeyer.
- Haberzettl, S. (2014). Zweitspracherwerb und Mehrsprachigkeit bei Kindern und Jugendlichen in der Migrationsgesellschaft. In S. Chilla & S. Haberzettl (Eds.): *Handbuch Spracherwerb und Sprachentwicklungsstörungen*. Mehrsprachigkeit. München: Elsevier/ Urban & Fischer, 3-18.
- Hackman, D. A., Betancourt, L. M., Gallop, R., Romer, D., Brodsky, N. L., Hurt, H., & Farah, M. J. (2014). Mapping the trajectory of socioeconomic disparity in working memory: parental and neighborhood factors. *Child Development*, 85(4), 1433-1445. <https://doi.org/10.1111/cdev.12242>
- Haddad, B., Awwad, A., Hattab, M., & Hattab, A. (2018). Associative Root-Pattern Data and Distribution in Arabic Morphology. *Data*, 3(2), 10. <https://doi.org/10.3390/data3020010>
- Haddad, R. (2022). *Child bilingualism in Sweden and Lebanon: A study of Arabic-speaking 4-to-7-year-olds*. [Doctoral dissertation, Uppsala University]. Uppsala: Acta Universitatis Upsaliensis.
- Haddad-Hanna, M., & Friedmann, N. (2014). On the acquisition of relative clauses and Wh questions in Palestinian Arabic speaking children. In G. Russo-Zimet, M. Ziv & A. Masarwa (Eds.) *Young Children in the Arab Society in Israel*. Tel Aviv: The Mofet Institute (in Hebrew).
- Haegeman, L. (1994). *Introduction to government and binding theory* (2nd ed.). Cambridge, MA: Blackwell.
- Haegeman, L. (1995). Root infinitives, tense and truncated structures in Dutch. *Language Acquisition*, 4, 205-255.
- Haegeman, L. (2012). *Adverbial clauses, main clause phenomena and the composition of the left periphery*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199858774.001.0001>
- Haffner, U. (1995). *Gut reden kann ich*. Dortmund: Verlag Modernes Lernen.
- Hager, M., & Müller, N. (2015). Ultimate attainment in bilingual first language acquisition. *Lingua* 164,

- 289-308. <https://doi.org/10.1016/j.lingua.2015.05.014>
- Haider, H. (1997). Projective economy- on the minimal function structure of German. In W. Abraham & E. van Gelderen (Eds.), *German: Syntactic problems – problematic syntax* (pp. 83-103). Tübingen: Niemeyer.
- Haider, H. (2010). *The syntax of German*. Cambridge: Cambridge University Press.
- Hair Jr., J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3rd ed.). New York: Macmillan.
- Håkansson, G. (2017). Typological and developmental considerations on specific language impairment in monolingual and bilingual children: A Processability Theory account. *Language Acquisition*, 24(3), 265-280. <https://doi.org/10.1080/10489223.2016.1192634>
- Håkansson, G., & Nettelbladt, U. (1993). Developmental sequences in L1 (normal and impaired) and L2 acquisition of Swedish. *International Journal of Applied Linguistics*, 3, 131-157.
- Håkansson, G., Salameh, E. K., & Nettelbladt, U. (2003). Measuring language development in bilingual children: Swedish-Arabic children with and without language impairment. *Linguistics*, 41(2), 255-288.
- Hall, T.A. (2001). The status of extrasyllabic consonants in English and German. *ZAS Papers in Linguistics (ZASPiL)*, 21, 89-117. <https://doi.org/10.21248/zaspil.21.2001.91>
- Hamann, C. (2000). Parameters and (L2) Acquisition: Verb-Raising. In *GG@G Generative Grammar in Geneva, I*, 275-291.
- Hamann, C. (2002). *From syntax to discourse: Pronominal clitics, null subjects and infinitives in child language*. Dordrecht: Kluwer Academic Publishers.
- Hamann, C. (2003). Phenomena in French normal and impaired language acquisition and their implications for hypotheses on language development. *Probus*, 15, 91-122.
- Hamann, C. (2006). Speculations about early syntax: The production of wh-questions by normally developing French children and French children with SLI. *Catalan Journal of Linguistics*, 5, 143-189.
- Hamann, C. (2012). Bilingual development and language assessment. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of BUCLD 36* (pp. 1-28). Somerville, MA: Cascadilla Press.
- Hamann, C. (2015). Specific language impairment in German speaking children. In S. Stavrakaki (Ed.), *Specific Language Impairment: Current Trends in Research* (pp. 215-251). Amsterdam: Benjamins. <https://doi.org/10.1075/lald.58.10ham>
- Hamann, C. (2017). Phonological working memory and language development: What are the measures and what do they measure? *Applied Psycholinguistics*, 38(6), 1313-1321. <https://doi.org/10.1017/S0142716417000261>
- Hamann, C., & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, 2, 1-19. <https://doi.org/10.3389/fcomm.2017.00016>
- Hamann, C., & Belletti, A. (2006). Developmental patterns in the acquisition of complement clitic pronouns: Comparing different acquisition modes with an emphasis on French. *Rivista di Grammatica Generativa*, 31, 39-78.
- Hamann, C., Chilla, S., Abed Ibrahim, L., & Fekete, I. (2020). Language assessment tools for Arabic-speaking heritage and refugee children in Germany. *Applied Psycholinguistics*, 41, 1375-1414. <https://doi.org/10.1017/S0142716420000399>
- Hamann, C., Chilla, S., Gagarina, N., & Abed Ibrahim, L. (2017). Syntactic complexity and bilingualism: how (a)typical bilinguals deal with complex structures. In E. Di Domenico (Ed.), *Complexity in Acquisition* (pp. 142-178). Cambridge: Scholars Publishing.
- Hamann, C., Chilla, S., Ruigendijk, E., & Abed Ibrahim, L. (2013). *A German Sentence Repetition Task: Testing Bilingual Russian/German Children*. Poster presented at the COST meeting in Krakow, May 2013. DOI 10.17605/OSF.IO/8KB4N
- Hamann, C., Lindner, K., & Penner, Z. (2001). Tense, reference time, and language impairment in German children. *Audiatu vox sapientiae. Festschrift für Arnim von Stechow*. Berlin: Akademie Verlag, 182-

213.

- Hamann, C., Ohayon, S., Dubé, S., Frauenfelder, U., Rizzi, L., Starke, M., & Zesiger, P. (2003). Aspects of grammatical development in young French children with SLI. *Developmental Science*, 6(2), 151-158. <https://doi.org/10.1111/1467-7687.00265>
- Hamann, C., Penner, Z., & Linder, K. (1998). German Impaired Grammar: The Clause Structure Revisited. *Language Acquisition*, 7(2-4), 193-245. https://doi.org/10.1207/s15327817la0702-4_5
- Hamann, C., & Plunkett, K. (1998). Subjectless sentences in child Danish. *Cognition*, 69(1), 35-72. [https://doi.org/10.1016/S0010-0277\(98\)00059-6](https://doi.org/10.1016/S0010-0277(98)00059-6)
- Hamann, C., Rizzi, L., & Frauenfelder, U.H. (1996). On the acquisition of subject and object clitics in French. In H. Clahsen (Ed.), *Generative perspectives on language acquisition* (pp. 309-334). Amsterdam: John Benjamins.
- Hamann, C., & Tuller, L. (2014). Genuine versus superficial relatives in French: the depth of embedding factor. *Revisita di Grammatica Generativa* 36, 146-181.
- Hamann, C., & Tuller, L. (2015). Intervention effects in the spontaneous production of relative clauses in (a)typical language development of French children and adolescents. In E. Di Domenico, C. Hamann & S. Matteini (Eds.), *Structures, Strategies and Beyond* (pp. 321-342). Amsterdam: John Benjamins. <https://doi.org/10.1075/la.223.15ham>
- Hamann, C., Tuller, L., Monjauze, C., Delage, H., & Henry, C. (2007). (Un)successful subordination in French-speaking children and adolescents with SLI. In H. Cauntulton, S. Kalatilake & I. Woo (Eds.), *Proceedings of BUCLD 31* (pp. 286-297). Somerville: Cascadia Press.
- Hambly, H., Wren, Y., McLeod, S., & Roulstone, S. (2013). The influence of bilingualism on speech production: a systematic review. *International Journal of Language & Communication Disorders*, 48(1), 1-24. <https://doi.org/10.1111/j.1460-6984.2012.00178.x>
- Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2005). The role of working memory in higher-level cognition. In R. Sternberg & J. E. Pretz (Eds.), *Cognition and Intelligence: Identifying the Mechanisms of the Mind* (pp. 104-121). New York, NY: Cambridge University Press.
- Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, 29, 715-733. <https://doi.org/10.1016/j.ecresq.2014.05.008>
- Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008). Exposure to English before and after entry into head start: Bilingual children's receptive language growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism*, 11(1), 30-56. <https://doi.org/10.2167/beb376.0>
- Hammond, M. (2003). *The phonology of English: A prosodic optimality-theoretic approach*. Oxford University Press.
- Hanke, M. (2014). *Strategies and Mechanisms for Language Production in Noise*. Erlensee: Selbstverlag.
- Hansen, D. (1994). Zur Wirksamkeit und Effizienz einer psycholinguistisch begründeten Methode der Sprachtherapie bei kindlichem Dysgrammatismus. *Sprache-Stimme-Gehör*, 18, 25-33.
- Hargreaves, I. S., Pexman, P. M., Johnson, J. C., & Zdrzilova, L. (2012). Richer concepts are better remembered: Number of features effects in free recall. *Frontiers in Human Neuroscience*, 6, 73. <https://doi.org/10.3389/fnhum.2012.00073>
- Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, 19, 379-400.
- Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H Brookes Publishing.
- Hartman, A., & Jackisch, A. (2013). Hamburger Bevölkerung mit Migrationshintergrund. *Statistische Analysen*, 6. Statistisches Amt für Hamburg und Schleswig-Holstein.
- Hartmann, J. (2016). Do second-generation Turkish migrants in Germany assimilate into the middle class? *Ethnicities*, 16(3), 368-392. <https://doi.org/10.1177/1468796814548234>
- Hatzidaki, A., Branigan, H. P., & Pickering, M. J. (2011). Co-activation of syntax in bilingual language

- production. *Cognitive Psychology*, 62(2), 123-150. <https://doi.org/10.1016/j.cogpsych.2010.10.002>
- Hawkins, R. (2001). The theoretical significance of universal grammar in SLA. *Second Language Research*, 17, 345-367.
- Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research*, 22(3), 269-301. <https://doi.org/10.1191/0267658306sr269oa>
- Haznedar, B. (1995). *Acquisition of English by a Turkish child: On the development of VP and negation. Paper presented at Language Acquisition Research Symposium (LARS)*, University of Utrecht, 11 May (Manuscript, University of Durham).
- Haznedar, B. (1997a). L2 acquisition by a Turkish-speaking child: Evidence for L1 influence. In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 245-256). Somerville, MA: Cascadilla Press.
- Haznedar, B. (1997b). *Child second language acquisition of English: a longitudinal case study of a Turkish-speaking child* (Doctoral dissertation, Durham University).
- Haznedar, B. (2001). The acquisition of the IP system in child L2 English. *Studies in Second Language Acquisition*, 23, 1-39. <https://doi.org/10.1017/S0272263101001012>
- Haznedar, B. (2003). The status of functional categories in child second language acquisition: Evidence from the acquisition of CP. *Second Language Research*, 19(1), 1-41. <https://doi.org/10.1191/0267658303sr212oa>
- Haznedar, B., & Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes and A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 257-268.
- Heckt, M., & Pohlmann, B. (2018). *Das Verfahren zur Vorstellung Viereinhalbjähriger. Ergebnisse Schuljahr 2017/18. Hamburg*. <https://ifbq.hamburg.de/wp-content/uploads/sites/803/2024/03/pdf-bericht-viereinhalbjaehrigenvorstellung-schuljahr-2017-2018.pdf>
- Hendriks, P., & Koster, C. (2010). Production /comprehension asymmetries in language acquisition. *Lingua*, 120(8), 1887-1897. <https://doi.org/10.1016/j.lingua.2010.02.002>
- Henry, G., & Saad, S. (2012). *Plurilinguisme et trouble spécifique du langage: Etude pilote sur la pertinence d'une épreuve de répétition de phrases au Liban*. [Master's thesis, Université Saint Joseph, Beirut].
- Henry, L.A. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, 9, 233-247.
- Henry, L.A. (2012). *The Development of Working Memory in Children*. London: Sage Publications.
- Henry, L. A., & Botting, N. (2017). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, 33(1), 19-32.
- Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of child psychology and psychiatry*, 53(1), 37-45.
- Herschensohn J (2007). *Language Development and Age*. Cambridge: Cambridge University Press.
- Hertel, I., Chilla, S., & Abed Ibrahim, L. (2022). Special Needs Assessment in Bilingual School-Age Children in Germany. *Languages*, 7(1), 4. <https://doi.org/10.3390/languages7010004>
- Hilchey, M., & Klein, R. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin and Review*, 18(4), 625-658. <https://doi.org/10.3758/s13423-011-0116-7>
- Hilchey, M. O., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 586-613). <https://doi.org/10.1017/CBO9781107447257.026>
- Hoff, E. (2003a). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child development*, 74(5), 1368-1378. <https://doi.org/10.1111/1467->

8624.00612

- Hoff, E. (2003b). Causes and consequences of SES-related differences in parent-to-child speech. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 147–160). Lawrence Erlbaum Associates Publishers.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55–88. <https://doi.org/10.1016/j.dr.2005.11.002>
- Hoff, E. (2018). Bilingual Development in Children of Immigrant Families. *Child development perspectives*, 12(2), 80–86. <https://doi.org/10.1111/cdep.12262>
- Hoff, E., Brett L., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of Parenting*, 2nd ed (pp. 231–52). Mahwah: Erlbaum.
- Hoff, E., Burrridge, A., Ribot, K. M., & Giguere, D. (2018a). Language specificity in the relation of maternal education to bilingual children's vocabulary growth. *Developmental psychology*, 54(6), 1011–1019. <https://doi.org/10.1037/dev0000492>
- Hoff, E., Quinn, J. M., & Giguere, D. (2018b). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, 21(2), 10.1111/desc.12536. <https://doi.org/10.1111/desc.12536>
- Hoff, E., & Core, C. (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, 34(4), 215–26. <https://doi.org/10.1055/s-0033-1353448>
- Hoff, E., Core, C., & Bridges, K. (2008). Nonword repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, 35(4), 903–916. <https://doi.org/10.1017/S0305000908008751>
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. <https://doi.org/10.1017/S0305000910000759>
- Hoff, E., Rumiche, R., Burrridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, 29(4), 433–444. <https://doi.org/10.1016/j.ecresq.2014.04.012>
- Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, 38(4), 271–278. <https://doi.org/10.1016/j.jcomdis.2005.02.003>
- Hoff, E., Quinn, J. M., & Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, 21(2), 10.1111/desc.12536. <https://doi.org/10.1111/desc.12536>
- Hole, C. (1995). *Modern Arabic: Structures, Functions, and Varieties*. New York, USA: Longman Linguistic Library.
- Holmberg, A. (2015). Verb second. In T. Kiss & Ar. Alexiadou (Eds.), *Syntax – theory and analysis: An international handbook* (pp. 342–382). Berlin & New York: Walter de Gruyter.
- Hoover, J., Storkel, H., & Rice, M. (2012). The interface between neighborhood density and optional infinitives: Normal development and specific language impairment. *Journal of Child Language*, 39, 835–62.
- Hopkins, B., and Skellam, J. G. (1954). A new method for determining the type of distribution of plant individuals. *Annals of Botany*, 18(2), 213–227. <https://doi.org/10.1093/oxfordjournals.aob.a083391>
- Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12(4), 463–483. <https://doi.org/10.1017/S1366728909990253>
- Hopp, H., & Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, 34(2), 361–394. <https://doi.org/10.1017/S0142716411000737>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference

- framework. *Journal of Computational Graphical Statistics*, 15(3), 651-674. <https://doi.org/10.1198/106186006X133933>
- Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *Test of Early Language Development (TELD)* (Third Edition). PRO-ED
- Hulk, A. (1996). The syntax of WH-questions in Child French. In F. Wijnen & W. Philip (Eds.), *Connecting Children's Language and Linguistic Theory. Amsterdam Series in Child Language Development*, 5 (pp. 129-172). Amsterdam.
- Hulk, A. (2000). L'acquisition des pronoms clitiques français par un enfant bilingue françaisnéerlandais. *The Canadian Journal of Linguistics* 45(1-2), 97-117. <https://doi.org/10.1017/S000841310001762X>
- Hulk, A., & Cornips, L. (2006). Between 2L1- and child L2 acquisition: An experimental study of bilingual Dutch. In C. Lleó (Ed.), *Interfaces in Multilingualism: Acquisition and representation* (pp. 115-137). Amsterdam: John Benjamins. <https://doi.org/10.1075/hsm.4.06hul>
- Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), 227-244. <https://doi.org/10.1017/S1366728900000353>
- Hulk, A., & Zuckerman, S. (2000). The interaction between input and economy: Acquiring optionality in French wh-questions. In S.C. Howel, S. A. Fish & T. Keith-Lucas (Eds.), *Proceedings of BUCLD 24* (pp. 438-449). Somerville: Cascadilla Press.
- Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology*, 32, 347-355.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6), F31-F39.
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61, 343-65.
- Hyams, N. M. (1986). *Language Acquisition and the Theory of Parameters*. Boston: Reidel.
- IBM Corp. (2016). *IBM SPSS Statistics for Windows*, Version 24.0. Armonk, NY: Author.
- IBM Corp. (2019). *IBM SPSS Statistics for Windows*, Version 26.0. Armonk, NY: IBM Corp
- IBM Corp. (2020). *IBM SPSS Statistics for Windows*, Version 27.0. Armonk, NY: IBM Corp
- Iluz-Cohen, P., & Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism: Language and Cognition*, 16(4), 884-899. <https://doi.org/10.1017/S1366728912000788>
- Ingham, R., Fletcher, P., Schelletter, C., & Sinka, I. (1998). Resultative VPs and specific language impairment. *Language Acquisition*, 7, 87-111.
- International Association of Logopedics and Phoniatrics. (2011). *Recommendations for working with bilingual children*. http://www.specchioriflesso.net/media/162083/linee_guida_bilingui_ialp_may_2011.pdf
- Ionin, T., Zubizarreta, M., & Philippov, V. (2009). Acquisition of article semantics by child and adult L2-English learners. *Bilingualism: Language and Cognition*, 12(3), 337-361. <https://doi.org/10.1017/S1366728909990149>
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127(2), 277-288. <https://doi.org/10.1016/j.actpsy.2007.06.003>
- Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language & Communication Disorders*, 51(1), 61-73.
- Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2020). Working, declarative, and procedural memory in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(12), 4162-4178.

- Jackson, C., Schatschneider, C., & Leacox, L. (2014). Longitudinal analysis of receptive vocabulary growth in young Spanish-English-speaking children from immigrant families. *Language, Speech and Hearing Services in Schools, 45*, 40-51.
- Jakubowicz, C. (2004). Is movement costly?. *Proceedings of the JEL 2004*, 33-34.
- Jakubowicz, C. (2005). The language faculty: (Ab)normal development and interface constraints. *Paper Presented at Generative Approaches to Language Acquisition, GALA, 2005*, Siena.
- Jakubowicz, C. (2011). Measuring derivational complexity: new evidence from typically developing and SLI learners of L1 French. *Lingua, 121*(3), 339-351. <https://doi.org/10.1016/j.lingua.2010.10.006>
- Jakubowicz, C., & Gutierrez, J. (2007). *Elicited production and comprehension of root wh questions in French and Basque. Presentation at the COST Meeting Cross Linguistically Robust Stage of Children's Linguistic Performance*. Berlin, Feb. 16th.
- Jakubowicz, C., Müller, N., Kang, O.K., Riemer, B., & Rigaut, C. (1996). On the acquisition of the pronominal system in French and German. In A. Springfellow, D. Cahana-Amitay, E. Hughes & A. Zukowski (Eds.), *Proceedings of BUCLD 20* (pp. 374-385). Somerville: Cascadia Press.
- Jakubowicz, C., & Nash, L. (2001). Functional categories and syntactic operations in (Ab)normal language acquisition. *Brain and Language, 77*, 321-339.
- Jakubowicz, C., Nash, L., Rigaut, C., & Gérard, C.-L. (1998). Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition: A Journal of Developmental Linguistics, 7*(2-4), 113-160. https://doi.org/10.1207/s15327817la0702-4_3
- Jakubowicz, C., & Rigaut, C. (2000). L'acquisition des clitiques nominatifs en français. *Revue Canadienne de Linguistique, Special issue on Language Acquisition, 45*, 119--157.
- Jakubowicz, C., & Tuller, L. (2008). Specific language impairment in French. In D. Ayoun (Ed.), *Studies in French applied linguistics* (pp. 7-134). Amsterdam: John Benjamins.
- Jakubowicz, C., & Strik, N. (2008). Scope-marking strategies in the acquisition of long distance wh-questions in French and Dutch. *Language and Speech, 51*(1-2), 101-132. <https://doi.org/10.1177/00238309080510010701>
- James, D., Van Steenbrugge, W., & Chiveralls, K. (1994). Underlying deficits in language-disordered children with central auditory processing difficulties. *Applied Psycholinguistics, 15*, 311-328.
- Jarrah, M. (2017). A Criterial Freezing approach to subject extraction in Jordanian Arabic. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 62*(3), 411-448.
- Jeffries, E., Lambon Ralph, M. A., & Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language, 51*, 623-643. <https://doi.org/10.1016/j.jml.2004.07.005>
- Jensen De López, K., Sundahl Olsen, L., & Chondrogianni, V. (2014). Annoying Danish relatives: Comprehension and production of relative clauses by Danish children with and without SLI. *Journal of Child Language, 41*(2014), 51-83. <https://doi.org/10.1017/S0305000912000517>
- Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics, 24*(1), 131-161. <https://doi.org/10.1017/S0142716403000079>
- Jia, G., & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: age-related differences. *Journal of Speech, Language, and Hearing Research, 50*(5), 1280-1299. [https://doi.org/10.1044/1092-4388\(2007/090\)](https://doi.org/10.1044/1092-4388(2007/090))
- Jia, R., & Paradis, J. (2020). The acquisition of Mandarin relative clauses by Mandarin heritage language children. *Linguistic Approaches to Bilingualism, 10*(2), 153-183. <https://doi.org/10.1075/lab.16015.jia>
- Johnson, C. J., Beitchman, J., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology, 19*(1), 51-65.
- Johnson, C. J., Beitchman, J., Young, A., Escobar, M., Atkinson, L., Wilson, B., ... & Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language

- stability and outcomes. *Journal of Speech, Language, and Hearing Research*, 42(3), 744-760.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60-99. [https://doi.org/10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0)
- Johnston, J. R., Smith, L. B., & Box, P. (1997). Cognition and communication: referential strategies used by preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 40(5), 964-974. <https://doi.org/10.1044/jslhr.4005.964>
- Jones, A., Atkinson, J., Marshall, C., Botting, N., St Clair, M. C., & Morgan, G. (2020). Expressive vocabulary predicts nonverbal executive function: A 2-year longitudinal study of deaf and hearing Children. *Child Development*, 91(2), e400-e414. <https://doi.org/10.1111/cdev.13226>
- Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language and Hearing Research*. 53(6), 1642-1655. [https://doi.org/10.1044/1092-4388\(2010/09-0222\)](https://doi.org/10.1044/1092-4388(2010/09-0222))
- Jongman, S. R., Roelofs, A., Scheper, A. R., & Meyer, A. S. (2016). Picture naming in typically developing and language-impaired children: the role of sustained attention. *International Journal of Language and Communication Disorders*, 52(3), 323-333. <https://doi.org/10.1111/1460-6984.12275>
- Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44, 137-166.
- Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech & Hearing Research*, 37(2), 418-421. <https://doi.org/10.1044/jslhr.3702.418>
- Kail, R., Hale, C. A., Leonard, L. B., & Nippold, M. A. (1984). Lexical storage and retrieval in language-impaired children. *Applied Psycholinguistics*, 5(1), 37-49.
- Kail, R., & Leonard, L. (1986). Word-finding abilities in language impaired children. *ASHA Monographs*, 25, 1-39.
- Kalashnikova, M., & Mattock, K. (2014). Maturation of executive functioning skills in early sequential bilingualism. *International Journal of Bilingual Education and Bilingualism*, 17(1), 111-123.
- Kalia V., Wilbourn M., & Ghio K. (2014). Better early or late? Examining the influence of age of exposure and language proficiency on executive function in early and late bilinguals. *Journal of Cognitive Psychology*, 26, 699-713.
- Kalvestinova, E. (2007). Three stages of root infinitive production in early child Russian. *First language*, 27(2), 99-131.
- Kalnak, N., Peyrard-Janvid, M., Forssberg, H., & Sahlén, B. (2014). Nonword repetition – A clinical marker for specific language impairment in Swedish associated with parents' language-related problems. *PLOS ONE*, 9(2), article e89544.
- Kaltenbacher, E., & Lindner, K. (1990). Some aspects of delayed and deviant development in language impaired children learning German. *Paper presented at the Meeting of the European Group for Child Language Disorders*, Røros, Norway.
- Kaltsa, M., Tsimpli, I. M., & Rothman, J. (2015). Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua*, 164, 266-288. <https://doi.org/10.1016/j.lingua.2015.06.002>
- Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2014). On the nature of verb-noun dissociations in bilingual SLI: A psycholinguistic perspective from Greek. *Bilingualism: Language and Cognition*, 17(1), 169-188.
- Kambanaros, M., Michaelides, M., & Grohmann, K. K. (2015). Measuring word retrieval deficits in a multilingual child with SLI: Is there a better language? *Journal of Neurolinguistics*, 34, 112-130.
- Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: lexical-semantic skills in L1 and L2. *Journal of Speech, Language, and Hearing Research*, 48(2), 372-383. [https://doi.org/10.1044/1092-4388\(2005/026\)](https://doi.org/10.1044/1092-4388(2005/026))

- Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53(3), 739-756. [https://doi.org/10.1044/1092-4388\(2009/08-0248\)](https://doi.org/10.1044/1092-4388(2009/08-0248))
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189-217. <https://doi.org/10.1037/0096-3445.133.2.189>
- Kapa, L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, 76, 237-252. <https://doi.org/10.1016/j.jml.2014.07.004>
- Kapa, L. L., & Plante, E. (2015). Executive Function in SLI: Recent Advances and Future Directions. *Current Developmental Disorders Reports*, 2(3), 245-252. <https://doi.org/10.1007/s40474-015-0050-x>
- Kaplan, D. (1983). *Order of acquisition of morph-syntactic categories among Hebrew-speaking 2 to 3 year olds*. [Master's thesis, Tel-Aviv University].
- Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawing. *Cognition*, 34(1), 57-83. [https://doi.org/10.1016/0010-0277\(90\)90031-E](https://doi.org/10.1016/0010-0277(90)90031-E)
- Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2022). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition*, 25(1), 29-41. <https://doi.org/10.1017/S1366728921000390>
- Kashef, R., & Kamel, M. S. (2008). Efficient bisecting k-medoids and its application in gene expression analysis. In A. Campilho & M. Kamel (Eds.), *ICIAR 2008 Proceedings of the 5th International Conference on Image Analysis and Recognition, LNCS 5112* (pp. 423-434). Heidelberg: Springer. https://doi.org/10.1007/978-3-540-69812-8_42
- Kassambara, A., & Mundt, F. (2017). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses, R Package Version 1.0.5*. Available at: <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf> [accessed June 1, 2018].
- Katsos, N., Roqueta, C. A., Estevan, R. A., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, 119, 43-57.
- Kaufman, L., & Rousseeuw, P. J. (1987). Clustering by means of Medoids. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods* (pp. 405-416). Basel: Birkhäuser Basel.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Kauschke, C. (2012). *Kindlicher Spracherwerb im Deutschen*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110283891>
- Kauschke, C., Renner, L. F., & Domahs, U. (2017). Past participle formation in specific language impairment. *International Journal of Language & Communication Disorders*, 52(2), 168-183. <https://doi.org/10.1111/1460-6984.12255>
- Kaushanskaya, M., Gross, M., & Buac, M. (2014). Effects of classroom bilingualism on task shifting, verbal memory, and word learning in children. *Developmental Science*, 17, 564-583.
- Kaushanskaya, M., & Marian, V. (2007). Bilingual Language Processing and Interference in Bilinguals: Evidence From Eye Tracking and Picture Naming. *Language Learning*, 57(1), 119-163. <https://doi.org/10.1111/j.1467-9922.2007.00401.x>
- Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition*, 18(1), 27-28. <https://doi.org/10.1017/S1366728914000510>
- Kavitskaya, D., Babyonyshev, M., Walls, T., & Grigorenko, E. L. (2011). Investigating the effects of syllable complexity in Russian-speaking children with SLI. *Journal of Child Language*, 38(5), 979-998.
- Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, 10(2), 159-180. <https://doi.org/10.1080/02687039608248403>

- Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
- Kay-Raining Bird, E., Genesee, F., & Verhoeven, L. (2016). Bilingualism in children with developmental disorders: A narrative review. *Journal of Communication Disorders*, 63, 1-14. <https://doi.org/10.1016/j.jcomdis.2016.07.003>
- Keffala, B., Scarpino, S., Hammer, C. S., Rodriguez, B., Lopez, L., & Goldstein, B. (2020). Vocabulary and Phonological Abilities Affect Dual Language Learners' Consonant Production Accuracy Within and Across Languages: A Large-Scale Study of 3- to 6-Year-Old Spanish-English Dual Language Learners. *American Journal of Speech-Language Pathology*, 29(3), 1196-1211. https://doi.org/10.1044/2019_AJSLP-19-00145
- Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *International Journal of Bilingualism*, 6(3), 315-334. <https://doi.org/10.1177/13670069020060030601>
- Kehoe, M. (2015). Cross-linguistic interaction: A retrospective and prospective view. In E. Babatsouli & D. Ingram (Eds.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015* (pp.141-167). Chania: Institute of Monolingual and Bilingual Speech. URL: <http://ismbs.eu/publications>
- Kehoe, M., Friend, M., & Poulin-Dubois, D. (2021). Relations between phonological production, grammar and the lexicon in bilingual French-English children. *The International Journal of Bilingualism: Cross-disciplinary, Cross-linguistic Studies of Language Behavior*, 25(6), 1576-1596. <https://doi.org/10.1177/13670069211031987>
- Kehoe, M., & Girardier, C. (2020). What factors influence phonological production in French-speaking bilingual children, aged three to six years?. *Journal of Child Language*, 47(5), 945-981. <https://doi.org/10.1017/S0305000919000874>
- Kehoe, M., & Havy, M. (2019). Bilingual phonological acquisition: The influence of language-internal, language-external, and lexical factors. *Journal of Child Language*, 46(2), 292-333. <https://doi.org/10.1017/S0305000918000478>
- Kehoe, M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition*, 7(1), 71-88. <https://doi.org/10.1017/S1366728904001282>
- Khamis-Dakwar, R., & Khattab, G. (2014). Cultural and linguistic considerations in language assessment and intervention for Levantine Arabic speaking children. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse (CLD) Populations*, 21(3), 78-87. <https://doi.org/10.1044/cds21.3.78>
- Khamis-Dakwar, R., & Makhoul, B. (2014). Arabic diglossia and its implications for language learning disability assessment. In E. Saiegh-Haddad & M. Joshi (Eds.), *Arabic literacy handbook: Theoretical insights and practical challenges* (pp. 279-303). Springer.
- Khater, M. (2016). *The relationship between nonword repetition, root and pattern effects, and vocabulary in Gulf Arabic-speaking children* [Unpublished Doctoral thesis, City, University of London], London, United Kingdom.
- Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. In D. B. Parkinson & E. Benmamoun (Eds.), *Current Issues in Linguistic Theory: Vol. 230. Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics* (pp. 1-37). John Benjamins.
- Khomsí, A. (2001). *Evaluation du langage oral* (ELO). ECPA: Paris.
- Khomsí, A., Khomsí, J., Pasquet, F., & Parbeau-Guénó, A. (2007). *Bilan Informatisé de langage Oral au cycle 3 et au Collège (BILO-3C)*. Paris: Éditions du Centre de psychologie appliquée.
- Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, 33(3), 208-223. <https://doi.org/10.1097/TLD.0b013e31829d623e>
- Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, 22(6), 860-897.

- Kim, J. Y. (2020). Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition*, 23(2), 233-250. <https://doi.org/10.1017/S1366728918001220>
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15-47. [https://doi.org/10.1016/0010-0277\(72\)90028-5](https://doi.org/10.1016/0010-0277(72)90028-5)
- Klumper, S., & Korecky-Kröll, K. (2002). Nouns and verbs at the transition from pre-to protomorphology: A longitudinal case study on Austrian German. In M.D. Voeikova & W.U. Dressler (Eds.), *Pre-and protomorphology: Early phases of morphological development in nouns and verbs* (pp. 61-74). München: Lincom.
- Klein, W. (2006). On finiteness. In V. Van Geengoven (Ed.), *Semantics in acquisition* (pp. 245-272). Springer.
- Klein, W., & Perdue, C. (1997). The Basic Variety, or: Couldn't Natural Language be much Simpler? *Second Language Research*, 13, 301-347.
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, 18(1), 146-154. <https://doi.org/10.1111/desc.12202>
- Klieme, Eckhard et al. (2010): *PISA 2009: Bilanz nach einem Jahrzehnt*. Münster.
- Kohnert K. (2010). Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, 43(6), 456-473. <https://doi.org/10.1016/j.jcomdis.2010.02.002>
- Kohnert, K. (2013). *Language disorders in bilingual children and adults*. 2nd edition. San Diego, CA: Plural Publishing.
- Kohnert, K., Ebert, K. D., & Pham, G. T. (2020). *Language disorders in bilingual children and adults*. Plural Publishing.
- Kohnert, K., Windsor, J., & Ebert, K. D. (2009). Primary or "specific" language impairment and children learning a second language. *Brain and language*, 109(2-3), 101-111.
- Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice*, 21, 19-29. <https://doi.org/10.1111/j.1540-5826.2006.00204.x>
- Komeili, M., & Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clinical Linguistics & Phonetics*, 27(2), 152-162. <https://doi.org/10.3109/02699206.2012.751625>
- Köpke, B. (2004). Neurolinguistic aspects of attrition. *Journal of Neurolinguistics*, 17(1), 3-30.
- Köpke, B. (2007). Language attrition at the cross-roads of brain, mind, society. In B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (Eds.), *Language and Attrition: Theoretical perspectives* (pp. 9-37). Amsterdam: John Benjamins.
- Köpke, B., & Genevskaja-Hanke, D. (2018). First language attrition and dominance: Same same or different? *Frontiers in Psychology*, 9, 1963. <https://doi.org/10.3389/fpsyg.2018.01963>
- Köpke, B., & Schmid, M. S. (2004). First language attrition: The next phase. In M. Schmid, B. Köpke, M. Keijzer & L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 1-43). Amsterdam: John Benjamins.
- Kosmidis, I. (2018). *brglm2: Bias Reduction in Generalized Linear Models*. R Package Version 0.1.7. Available at: <https://github.com/ikosmidis/brglm2>
- Koster, J. (1975). Dutch as an SOV Language. *Linguistic Analysis*, 1, 111-36.
- Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, 106(16), 6556-6560.
- Krassowski, E., & Plante, E. (1997). IQ variability in children with SLI: implications for use of cognitive referencing in determining SLI. *Journal of Communication Disorders*, 30(1), 1-9. [https://doi.org/10.1016/0021-9924\(95\)00052-6](https://doi.org/10.1016/0021-9924(95)00052-6)

- Krishna, S., Yesu Babu, A., & Kiran Kumar, R. (2018). Determination of optimal clusters for a Non-hierarchical clustering paradigm K-Means algorithm. In N. Devarakonda, N. Chaki, & A. Cortesi (Eds.), *Proceedings of International Conference on Computational Intelligence and Data Engineering*. Singapore: Springer (pp. 301-3169). <https://doi.org/10.1007/978-981-10-6319-0>
- Kroffke, S., & Rothweiler, M. (2006). Variation im frühen Zweitspracherwerb des Deutschen durch Kinder mit türkischer Erstsprache. In M. Vliegen (Ed.), *Variation in Sprachtheorie und Spracherwerb. Akten des 39. Linguistischen Kolloquiums* (pp. 145-153). Amsterdam. – Frankfurt a.M.
- Kroffke, S., Rothweiler, M., & Babur, E. (2007). Turkish German successive bilingual children: Age of onset and the development of German V2. *Paper presented at the 6th International Symposium on Bilingualism*. Hamburg, May 2007.
- Krok, W., & Leonard, L. (2015). Past tense production in children with and without specific language impairment across Germanic languages: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 58(4), 1326-1340. https://doi.org/10.1044/2015_JSLHR-L-14-0348
- Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25(5), 497-514. <https://doi.org/10.1080/20445911.2013.799170>
- Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, 128(3), 416-430.
- Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, 1, 377-394. <https://doi.org/10.1146/annurev-linguist-030514-124937>
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Valdes Kroff, J. R. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. *Psychology of Learning and Motivation. Advances in Research and Theory*, 56, 229- 262.
- Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, 13(3), 373-381. <https://doi.org/10.1017/S136672891000009X>
- Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., Blair, C. B., & Family Life Project Key Investigators. (2016). The contribution of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *Journal of Experimental Child Psychology*, 148, 20-34. <https://doi.org/10.1016/j.jecp.2016.03.008>
- Kupisch, T. (2005). Acceleration in bilingual first language acquisition. In T. Gaerts & H. Jacobs (Eds.), *Romance languages and linguistic theory 2003: selected papers from 'Going Romance' 2003* (pp.143 -160). Amsterdam: John Benjamins.
- Kupisch, T. (2006). *The Acquisition of Determiners in Bilingual German-Italian and German-French Children*. München: Lingcom Europa.
- Kupisch, T. (2007). Testing the effects of frequency on the rate of learning: Determiner use in early French, German and Italian. In I. Gülzow & N. Gagarina (Ed.), *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept* (pp. 83-114). Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110977905.83>
- Kupisch, T. (2008). Dominance, mixing and cross-linguistic influence: on their relation in bilingual development. In P. GuijarroFuentes, P. Larrañaga, & J. Clibbens, *First Language Acquisition of Morphology and Syntax: Perspectives Across Languages and Learners*. Amsterdam: John Benjamins (pp. 209-234).
- Kupisch, T. (2019). 2L1 Simultaneous Bilinguals as Heritage Speakers. In M. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 458-469). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198793595.013.36>
- Kupisch, T., Bayram, F., & Rothman, J. (2017). Terminology matters II: Early bilinguals show cross-linguistic influence but are not attriters. *Linguistic Approaches to Bilingualism*, 7(6), 719-724. <https://doi.org/10.1075/lab.00012.kup>
- Kupisch, T., Kolb, N., Rodina, Y., & Urek, O. (2021). Foreign Accent in Pre- and Primary School Heritage Bilinguals. *Languages*, 6(2), 96. <https://doi.org/10.3390/languages6020096>

- Kupisch, T., Müller, N., & Cantone, K. F. (2002). Gender in Monolingual and Bilingual First Language Acquisition: Comparing Italian and French. *Lingue e Linguaggio*, 107-150. <https://doi.org/10.1418/7559>
- Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 22(5), 564-582. <https://doi.org/10.1177/1367006916654355>
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models*, 4th Edn. Chicago, IL: McGraw-Hill/ Irwin.
- Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.
- Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413(6855), 519-523.
- Laloi, A. (2015). *Language and executive functioning in the context of specific language impairment and bilingualism* [Doctoral thesis, University of Amsterdam].
- La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1118-33
- Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14, 359-375.
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 102-129). Malden, MA: Blackwell.
- Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl & H. Goodluck (Eds.), *The role of features in second language acquisition* (pp. 106-140). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25, 173-227. <https://doi.org/10.1177/026765830>
- Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *The Quarterly Journal of Experimental Psychology*, 71, 2207-2222. <https://doi.org/10.1177/1747021817739834>
- Lauro, J., Core, C., & Hoff, E. (2020). Explaining individual differences in trajectories of simultaneous bilingual development: Contributions of child and environmental factors. *Child Development*, 91, 2063-2082. <https://doi.org/10.1111/cdev.13409>
- Law, J., Reilly, S., & Snow, P. C. (2013). Child speech, language and communication need re-examined in a public health context: a new direction for the speech and language therapy profession. *International Journal of Language & Communication Disorders*, 48(5), 486-496.
- Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*, 52, 1401-1416. [https://doi.org/10.1044/1092-4388\(2009/08-0142\)](https://doi.org/10.1044/1092-4388(2009/08-0142))
- Law, J., Tomblin, J. B., & Zhang, X. (2008). Characterizing the growth trajectories of language-impaired children between 7 and 11 years of age. *Journal of Speech, Language, and Hearing Research*, 51(3), 739-749. [https://doi.org/10.1044/1092-4388\(2008/052\)](https://doi.org/10.1044/1092-4388(2008/052))
- Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology*, 46(2), 417-427. <https://doi.org/10.1037/a0018540>
- Leclercq, A. L., Maillart, C., & Majerus, S. (2013). Nonword repetition problems in children with specific language impairment: a deficit in accessing long-term linguistic representations? *Topics in Language Disorders*, 33(3), 238-254. <https://doi.org/10.1097/TLD.0b013e31829dd8c9>
- Leclercq, A. L., Quémart, P., Magis, D., & Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Research in Developmental Disabilities*, 35(12), 3423-3430. <https://doi.org/10.1016/j.ridd.2014.08.026>
- Lein, T., Hamann, C., Rothweiler, M., Abed Ibrahim, L., Chilla, S., & San, H. (2016). SLI in bilinguals:

- testing complex syntax and semantics in German. In D. Stringer, J. Garrett, B. Halloran, & S. Mossman (Eds.), *GASLA Proceedings 2015* (pp. 124-135). Somerville, MA: Cascadia Press.
- Lein, T., Rothweiler, M., & Hamann, C. (2017). Factors affecting the performance in child heritage Portuguese in Germany. In J. Choi, H. Demirdache, O. Lungu, & L. Voeltzel (Eds.), *Proceedings of GALA 2015* (pp.146-175). Newcastle: Cambridge Scholars Publishing.
- Leivada, E., Westergaard, M., Duñabeitia, J. A., & Rothman, J. (2021). On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed?. *Bilingualism: Language and Cognition*, 24(1), 197-210.
- Lemmer, R. (2018). *Sprachentwicklungsstörungen bei frühen Zweitsprachlernern – Der Erwerb von Kasus, Finitheit und Verbstellung* [Doctoral dissertation, Goethe-Universität Frankfurt]. urn:nbn:de:hebis:30:3-465833
- Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, 2(12), 59-67.
- Leonard, L. B. (1998). *Children with Specific Language Impairment*. Cambridge: MIT Press.
- Leonard, L. B. (2014a). *Children with Specific Language Impairment*, (2nd ed.). Cambridge: MIT Press.
- Leonard, L. B. (2014b). Specific language impairment across languages. *Child Development Perspectives*, 8, 1–5. <https://doi.org/10.1111/cdep.12053>
- Leonard, L. B. (2014c). Children with specific language impairment and their contribution to the study of language development. *Journal of Child Language* 41, 37–48. <https://doi.org/10.1017/S0305000914000130>
- Leonard, L. B. (2015). Language symptoms and their possible sources of Specific Language Impairment. In E. Bavin & L. Naigles (Eds.), *The Cambridge Handbook of Child Language* (Cambridge Handbooks in Language and Linguistics, pp. 545-563). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781316095829.025>
- Leonard, L. B., Camarata, S., Rowan, L., & Chapman, K. (1982). The communicative functions of lexical usage by language impaired children. *Applied Psycholinguistics*, 3, 109-125. <https://doi.org/10.1017/S0142716400006652>
- Leonard, L. B., Christ, S. L., Deevy, P., Karpicke, J. D., Weber, C., Haebig, E., Kueser, J. B., Souto, S., & Krok, W. (2021). A multi-study examination of the role of repeated spaced retrieval in the word learning of children with developmental language disorder. *Journal of Neurodevelopmental Disorders*, 13(1), 20. <https://doi.org/10.1186/s11689-021-09368-z>
- Leonard, L. B., & Deevy, P. (2004). Lexical Deficits in Specific Language Impairment. In L. Verhoeven, & H. van Balkom (Eds.), *Classification of Developmental Language Disorders-Theoretical Issues and Clinical Implications* (pp. 209-233). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Leonard, L. B., & Deevy, P. (2020). Retrieval practice and word learning in children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research*, 63(10), 3252-3262. https://doi.org/10.1044/2020_JSLHR-20-00006
- Leonard, L. B., & Dromi, E. (1994). The use of Hebrew verb morphology by children with specific language impairment and by children developing language normally. *First Language*, 14, 283-304.
- Leonard, L. B., Eyer, J., Bedore, L., & Grela, B. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 40(4), 741-753. <https://doi.org/10.1044/jslhr.4004.741>
- Leonard, L. B., Hansson, K., Nettelbladt, U., & Deevy, P. (2004). Specific language impairment in children: A comparison of English and Swedish. *Language Acquisition*, 12(3-4), 219-246. <https://doi.org/10.1080/10489223.1995.9671744>
- Leonard, L., & Kueser, J. (2019). Five overarching factors central to grammatical learning and treatment in children with developmental language disorder. *International Journal of Language & Communication Disorders*, 54(3), 347-361. <https://doi.org/10.1111/1460-6984.12456>
- Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42(3), 678–689. <https://doi.org/10.1044/jslhr.4203.678>

- Leonard, L. B., Weismer, S., Miller, C., Francis, D., Tomblin, J., & Kail, R. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research, 50*(2), 408-428. [https://doi.org/10.1044/1092-4388\(2007/029\)](https://doi.org/10.1044/1092-4388(2007/029))
- Levshina, N. (2015). *How to do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.195>
- Levy, R., E. Fedorenko, & T. Gibson. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language, 69*(4), 461-495.
- Levy, Y., & Kavé, G. (1999). Language breakdown and linguistic theory: a tutorial overview. *Lingua 107*(1-2), 95-143. [https://doi.org/10.1016/S0024-3841\(98\)00023-0](https://doi.org/10.1016/S0024-3841(98)00023-0)
- Lindner, K. (1995). The Acquisition and Use of Morphological Markers: What Is the Matter with SLI Children? In M. Kovačević (Ed.) *Language and Language Communication Barriers* (pp. 71-95). Zagreb: Croatian University Press.
- Lindner, K. (2002). Finiteness and children with specific language impairment: an exploratory study. *Linguistics 40*(4), 797-847. <https://doi.org/10.1515/ling.2002.031>
- Lindner, K. (2003). The development of sentence interpretation strategies in monolingual German-learning children with and without language impairment. *Special issue on specific language impairment. Linguistics, 41*, 213 -254.
- Liu, X., Tu, L., Chen, X., Wang, J., Li, M., Lu, Z., & Huang, R. (2021). Effect of AoA-L2 on L1 and L2 networks in early and late bilinguals. *International Journal of Bilingualism, 25*(6) 1-19. <https://doi.org/10.1177/13670069211033026>.
- Lleó, C. (2018). Aspects of the Phonology of Spanish as a Heritage Language: From Incomplete Acquisition to Transfer. *Bilingualism: Language and Cognition, 21*(4), 732-747. <https://doi.org/10.1017/S1366728917000165>
- Loaiza, V. M., & Camos, V. (2018). The role of semantic representations in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 863-881. <https://doi.org/10.1037/xlm0000475>
- Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: implications for the early years and beyond. *International Journal of Language and Communication Disorders, 37*(1), 3-15.
- Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language, 58*, 265-326.
- Loeb, D. F., Pye, C., Redmond, S., & Richardson, L. (1996). Eliciting verbs from children with specific language impairment. *American Journal of Speech-Language Pathology, 5*, 17-30.
- Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189-239). Academic Press.
- Logie, R. H. (1995). *Visuo-Spatial Working Memory*. Hove: Lawrence Erlbaum.
- Lohndal, T., Rothman, J., Kupisch, T., & Westergaard, M. (2019). Heritage language acquisition: What it reveals and why it is important for formal linguistic theories. *Language and Linguistics Compass, 13*(12), e12357. <https://doi.org/10.1111/lnc3.12357>
- Lombardi, L., & Potter, M.C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language, 31*(6), 713-733. [http://dx.doi.org/10.1016/0749-596X\(92\)90036-W](http://dx.doi.org/10.1016/0749-596X(92)90036-W)
- Long, M. H. (1990). maturational Constraints on Language Development. *Studies in Second Language Acquisition, 12*(3), 251-285. <https://doi.org/10.1017/S0272263100009165>
- Long, S. L. (1997). *Regression Models for Categorical and Limited Dependent Variables*. London: SAGE Publications.
- Lowe, C. J., Cho, I., Goldsmith, S. F., & Morton, J. B. (2021). The Bilingual Advantage in Children's Executive Functioning Is Not Related to Language Status: A Meta-Analytic Review. *Psychological Science, 32*(7), 1115-1146. <https://doi.org/10.1177/0956797621993108>
- Lu, C.H., & Proctor, R.W. (1995). The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects. *Psychonomic Bulletin and Review, 2*, 174-207.

- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.
- Luk, G. (2015). Who are the bilinguals (and monolinguals)? *Bilingualism: Language and Cognition, 18*, 35-36.
- Luk, G., Anderson, J. A., Craik, F. I., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and cognition, 74*(3), 347-357. <https://doi.org/10.1016/j.bandc.2010.09.004>
- Lum, J., & Bleses, D. (2012). Declarative and procedural memory in Danish-speaking children with specific language impairment. *Journal of Communication Disorders, 45*, 46-58.
- Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex, 51*, 1-10.
- Luo, L., Craik, F. I. M., Moreno, S., & Bialystok, E. (2013). Bilingualism interacts with domain in a working memory task: Evidence from aging. *Psychology and Aging, 28*(1), 28-34. <https://doi.org/10.1037/a0030875>
- Lustigman, L., & Berman, R. A. (2016). Form and function in early clause-combining. *Journal of Child Language, 43*(1), 157-185. <https://doi.org/10.1017/S0305000915000100>
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge, Cambridgeshire: Cambridge University Press.
- Maddieson, I. (2006). Correlating phonological complexity: data and validation. *Linguistic Typology 10*, 106-123.
- Maddieson, I., Flavier, S., Marsico, E., & Pellegrino, F. (2011). *LAPSyD: LyonAlbuquerque Phonological Systems Databases, Version 1.0*. Available at: <http://www.lapsyd.ddl.ish-lyon.cnrs.fr/>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2017). *Cluster: Cluster Analysis Basics and Extensions. R package version 2.0.1.2015*.
- Maguire, M. J., & Schneider, J. M. (2019). Socioeconomic status related differences in resting state EEG activity correspond to differences in vocabulary and working memory in grade school. *Brain and Cognition, 137*, 103619. <https://doi.org/10.1016/j.bandc.2019.103619>
- Maillart, C., Schelstraete, M. A., & Hupet, M. (2004). Phonological representations in children with SLI: a study of French. *Journal of Speech, Language, and Hearing Research, 47*(1), 187-198. [https://doi.org/10.1044/1092-4388\(2004\)016](https://doi.org/10.1044/1092-4388(2004)016)
- Mainela-Arnold, E., & Evans, J. L. (2005). Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research, 48*(4), 897-909. [https://doi.org/10.1044/1092-4388\(2005\)062](https://doi.org/10.1044/1092-4388(2005)062)
- Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*(2), 95-119. <https://doi.org/10.1016/j.jecp.2005.07.005>
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*, 50-68.
- Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in relative clause processing: The hiker that rocks crush. *Journal of Memory and Language, 54*: 466 - 490. <https://doi.org/10.1016/j.jml.2006.01.001>
- Mak, W. M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory and Cognition, 36*(1), 170-181. <https://doi.org/10.3758/MC.36.1.170>
- Makrodimitris, C., & Schulz, P. (2021). Does Timing in Acquisition Modulate Heritage Children's Language Abilities? Evidence from the Greek LITMUS Sentence Repetition Task. *Languages, 6*(1), 49. <https://doi.org/10.3390/languages6010049>
- Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish-English bilinguals. *Journal of Child Language, 37*(4), 817-840. <https://doi.org/10.1017/S0305000909990055>

- Marcovitch, S., & Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science*, 12(1), 1-18.
- Marinellie, S. A. (2004). Complex syntax used by school-age children with specific language impairment (SLI) in child-adult conversation. *Journal of Communication Disorders*, 37(6), 517-533.
- Marini, A., Eliseeva, N., & Fabbro, F. (2019). Impact of early second-language acquisition on the development of first language and verbal short-term and working memory. *International Journal of Bilingual Education and Bilingualism*, 22(2), 165-176. <https://doi.org/10.1080/13670050.2016.1238865>
- Marinis, T. (2011). On the nature and cause of specific language impairment: a view from sentence processing and infant research. *Lingua* 121(3), 463-475. <https://doi.org/10.1016/j.lingua.2010.10.010>
- Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95-124). Bristol: Multilingual Matters.
- Marinis, T., Armon-Lotem, S., & Pontikas, G. (2017). Language impairment in bilingual children - state of the art 2017. In J. Rothman & S. Unsworth (Eds.), *Linguistic Approaches to Bilingualism* (pp. 265-276). Amsterdam: John Benjamins.
- Marinis, T., & van der Lely, H. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *International Journal of Language and Communication Disorders*, 42(5), 557-582. <https://doi.org/10.1080/13682820601058190>
- Marinova-Todd, S. H. (2012). "Corplum is a core from a plum": The advantage of bilingual children in the analysis of word meaning. *Bilingualism: Language and Cognition*, 15(1), 117-127. <https://doi.org/10.1017/S136672891000043X>
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, 20(2), 121-157. [https://doi.org/10.1016/0010-0285\(88\)90017-5](https://doi.org/10.1016/0010-0285(88)90017-5)
- Marshall, C. R. (2014). Word production errors in children with developmental language impairments. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20120389. <https://doi.org/10.1098/rstb.2012.0389>.
- Marshall, C. R. (2020). Investigating the relationship between syntactic and short-term/working memory impairments in children with developmental disorders is not a straightforward endeavour. *First Language*, 40(4), 491-499.
- Marshall, C. R., Ebbels, S., Harris, J., & van der Lely, H. K. (2002). Investigating the impact of prosodic complexity on the speech of children with specific language impairment. *UCL Working Papers in Linguistics*. 14, 43-68.
- Marshall, C. R., Harris, J., & van der Lely, H. K. J. (2003). The nature of phonological representations in children with Grammatical-Specific Language Impairment (G-SLI). In Di. Hall, T. Markopoulos, A. Salamoura & S. Skoufaki (Eds.), *Proceedings of the University of Cambridge First Postgraduate Conference in Language Research* (pp. 511-517). Cambridge, UK: Cambridge Institute of Language Research
- Marshall, C. R., & Van der Lely, H. K. J. (2006). A challenge to current models of past tense inflection: The impact of phonotactics. *Cognition*, 100, 302-320.
- Marshall, C. R., & van der Lely, H. (2009). Effects of word position and stress on onset cluster production: evidence from typical development, SLI and dyslexia. *Language* 85(1), 39-57. <https://dx.doi.org/10.1353/lan.0.0081>.
- Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: language and cognition*, 11(1), 81-93.
- Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *International Journal of Language and Communication Disorders*, 43(2), 181-200.
- Marton, K. (2016). Executive control in bilingual children: Factors that influence the outcomes. *Linguistic Approaches to Bilingualism*, 6(5), 575-589.
- Marton, K., Campanelli, L., Eichorn, N., Scheuer, J., & Yoon, J. (2014). Information processing and proactive interference in children with and without specific language impairment. *Journal of Speech*,

- Language and Hearing Research*, 57(1), 106-119. [https://doi.org/10.1044/1092-4388\(2013/12-0306\)](https://doi.org/10.1044/1092-4388(2013/12-0306))
- Mateu, V. E. (2015). Object clitic omission in child Spanish: evaluating representational and processing accounts. *Language Acquisition*, 22, 240-284.
- Mateus, M.H. M., & d'Andrade Pardal, E. (2002). *The phonology of Portuguese*. Oxford: Oxford University Press.
- Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, 14(3), 186-223. <https://doi.org/10.1080/15475441.2018.1455584>.
- Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and developmental receptive language disorder—A comparative follow-up in early adult life. I: Cognitive and language outcomes. *Journal of Child Psychology and Psychiatry*, 41, 547-559.
- McCarthy, K., Mahon, M., Rosen, S., & Evans, B. (2014). Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition. *Child Development*, 85(5), 1965-1980. <https://doi.org/10.1111/cdev.12275>.
- McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research*, 60(10), 2891-2905.
- McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, 56(6), 1845-1856. [https://doi.org/10.1044/1092-4388\(2013/12-0233\)](https://doi.org/10.1044/1092-4388(2013/12-0233))
- McKean, C., Letts, C., & Howard, D. (2013). Developmental change is key to understanding primary language impairment: The case of phonotactic probability and nonword repetition. *Journal of Speech, Language, and Hearing Research*, 56(5), 1579-1594. [https://doi.org/10.1044/1092-4388\(2013/12-0066\)](https://doi.org/10.1044/1092-4388(2013/12-0066))
- McKinley, N., & Larson, V. (1989). Students who can't communicate: speech-language service at the secondary level. *Curriculum Report*, 19, 1-8.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53(2), 185-204. <https://doi.org/10.1037/0003-066X.53.2.185>.
- Meir, N. (2017). Effects of specific language impairment (SLI) and bilingualism on verbal short-term memory. *Linguistic Approaches to Bilingualism*, 7(3-4), 301-330. <https://doi.org/10.1075/lab.15033.mei>
- Meir, N. (2018). Morpho-syntactic abilities of unbalanced bilingual children: A closer look at the weaker language. *Frontiers in Psychology*, 9, 1318. <https://doi.org/10.3389/fpsyg.2018.01318>
- Meir, N., & Armon-Lotem, S. (2015). Disentangling bilingualism from SLI in Heritage Russian: the impact of L2 properties and length of exposure to the L2. In C. Hamman and E. Ruigendijk (Eds.), *Language Acquisition and Development: Proceedings of GALA 2013* (pp. 299-314). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, 8, 01442. <https://doi.org/10.3389/fpsyg.2017.01442>.
- Meir, N., Walters, J., & Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: the impact of L1 and L2 properties. *International Journal of Bilingualism*, 20(4), 421-452. <https://doi.org/10.1177/1367006915609240>
- Meir, N., Walters, J., & Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguistic Approaches to Bilingualism*, 7(5), 514-553. <https://doi.org/10.1075/lab.15007.mei>.
- Meisel, J. M. (1994). *Bilingual first language acquisition: French and German grammatical development*. Amsterdam & Philadelphia, PA: John Benjamins.
- Meisel, J. M. (2004). The bilingual child. In T.K. Bhatia & W.C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 91- 113). Oxford: Blackwell

- Meisel, J. M. (2007). The weaker language in early child bilingualism: Acquiring a first language as a second language? *Applied Psycholinguistics*, 28(3), 495-514. <https://doi.org/10.1017/S0142716407070270>
- Meisel, J. M. (2008). Child second language acquisition or successive first language acquisition? In B. Haznedar & E. Gavrusseva (Eds.), *Current trends in child second language acquisition: A generative perspective* (pp. 55-80). Amsterdam: John Benjamins. <https://doi.org/10.1075/lald.46.04mei>
- Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, 28(1), 5-34. <https://doi.org/10.1515/ZFSW.2009.002>
- Meisel, J. M. (2011). *First and second language acquisition: Parallels and differences*. Cambridge: Cambridge University Press.
- Meisel, J. M. (2013). Sensitive phases in successive language acquisition: The critical period hypothesis revisited. In Boeckx, C., & Grohmann, K. (Eds.), *The Cambridge handbook of biolinguistics* (pp. 69-85). Cambridge University Press.
- Meisel, J. M. (2018). Early child second language acquisition: French gender in German children. *Bilingualism: Language and Cognition*, 21(4), 656-673. <https://doi.org/10.1017/S1366728916000237>
- Meisel, J. M., Elsig, M., & Bonnesen, M. (2011). Delayed grammatical acquisition in first language development: subject-verb inversion and subject clitics in French interrogatives. *Linguistic Approaches to Bilingualism* 1(4), 347-390.
- Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, 23(10), 1092-1098. <https://doi.org/10.1177/0956797612443833>
- Melzer, J., Ring, A., Petermann, F., & Rißling, J.-K. (2018). Phonemerwerb monolingualer und mehrsprachiger Kinder im Vorschulalter. *Logos*, 26(2), 84-92.
- Messer, M. (2010). *Verbal short-term memory and vocabulary development in monolingual Dutch and bilingual Turkish-Dutch preschoolers* [Doctoral dissertation, Utrecht University].
- Messer, D., & Dockrell, J. E. (2013). Children with word finding difficulties: Continuities and profiles of abilities. *First Language*, 33(5), 433-448. <https://doi.org/10.1177/0142723713493345>
- Messer, M., Leseman, P. P., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, 105(4), 306-323. <https://doi.org/10.1016/j.jecp.2009.12.006>
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91(1), 3-19. <https://doi.org/10.1037/0022-0663.91.1.3>
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L.C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89-120). Mahwah, NJ: Erlbaum.
- Metz, M., van Hout, A., & van der Lely, H., (2010). Understanding who and which questions in five to nine-year-old Dutch children: The role of number. *New developments in the acquisition of Dutch, Groninger Arbeiten zur germanischen Linguistik*, 51, 27-41. <https://ugp.rug.nl/GAGL/article/view/30515>
- Miller, C. A., Leonard, L. B., & Finneran, D. (2008). Grammaticality judgements in adolescents with and without language impairment. *International Journal of Language & Communication Disorders*, 43(3), 346-360. <https://doi.org/10.1080/13682820701546813>
- Mills, A. (1985). The acquisition of German. In D. Slobin (Ed.). *The Cross-linguistic Study of Language Acquisition* (vol. 1, pp. 141-254). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mishra, R., & Abutalebi, J. (2020). Introduction to the special issue on cognitive consequences of bilingualism. *Journal of Cultural Cognitive Science*, 4(2), 123-125. <https://doi.org/10.1007/s41809-020-00070-5>
- Misyak, J., & Christiansen, M. (2011). Genetic variation and individual differences in language. In I. Arnon & E. V. Clark (Eds.), *Experience, variation, and generalization: learning a first language (Trends in*

- Language Acquisition Research*, vol. 7) (pp. 223-238). Amsterdam: John Benjamins.
- Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Erlbaum.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8-14. <https://doi.org/10.1177/0963721411429458>.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, 41(1), 49-100. <https://doi.org/10.1006/cogp.1999.0734>
- Moll, K., Hulme, C., Nag, S., & Snowling, M. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Applied Psycholinguistics*, 36(2), 203-221. <https://doi.org/10.1017/S0142716413000209>
- Montanari, E., & Akıncı, M., & Abel, R. (2019). Balance and dominance in the vocabulary of German-Turkish primary schoolchildren. *European Journal of Applied Linguistics*, 7(1), 113-144. <https://doi.org/10.1515/eujal-2018-0003>.
- Montanari, S., Mayr, R., & Subrahmanyam, K. (2018). Bilingual Speech Sound Development During the Preschool Years: The Role of Language Proficiency and Cross-Linguistic Relatedness. *Journal of Speech, Language, and Hearing Research*, 61(10), 2467-2486. https://doi.org/10.1044/2018_JSLHR-S-17-0393
- Montgomery J. W. (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics*, 16, 355-378.
- Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology*, 11(1), 77-91.
- Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: Effects of input rate and phonological working memory. *International Journal of Language & Communication Disorders*, 39(1), 115-133.
- Montgomery, J. W. (2008). Role of auditory attention in the real-time processing of simple grammar by children with specific language impairment: a preliminary investigation. *International Journal of Language and Communication Disorders*, 43(5), 499-527. <https://doi.org/10.1080/13682820701736638>.
- Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 269-288.
- Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., & Gillam, R. B. (2018a). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 61(12), 2950-2976. https://doi.org/10.1044/2018_JSLHR-L-17-0421
- Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: a preliminary study. *Applied Psycholinguistics*, 30(1), 123-151. <https://doi.org/10.1017/S0142716408090061>.
- Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2018b). Memory and language in children with SLI. In T.P. Alloway, *Working memory and clinical developmental disorders: Theories, debates and interventions* (pp. 22-37). London: Routledge.
- Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2016). Syntactic versus memory accounts of the sentence comprehension deficits of specific language impairment: Looking back, looking ahead. *Journal of Speech, Language, and Hearing Research*, 59(6), 1491-1504.
- Montgomery, J. W., Gillam, R. B., Evans, J. L., Schwartz, S., & Fargo, J. D. (2019). A comparison of the storage-only deficit and joint mechanism deficit hypotheses of the verbal working memory storage capacity limitation of children with developmental language disorder. *Journal of Speech, Language,*

- and *Hearing Research*, 62(10), 3808-3825. https://doi.org/10.1044/2019_JSLHR-L-19-0071
- Montgomery, J.W., Magimairaj, B.M., & Finney, M.C. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, 19, 78-94.
- Montgomery, J. W., Magimairaj, B. M., & O'Malley, M. H. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of psycholinguistic research*, 37, 331-354.
- Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.
- Montrul, S. (2016a). *The acquisition of heritage languages*. Cambridge University Press.
- Montrul, S. (2016b). Age of onset of bilingualism effects and availability of input in first language attrition. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 141-161). American Psychological Association. <https://doi.org/10.1037/14939-009>
- Montrul, S. (2018). Heritage language development: Connecting the dots. *International Journal of Bilingualism*, 22(5), 530-546. <https://doi.org/10.1177/1367006916654>
- Montrul, S. (2022). *Native Speakers, Interrupted: Differential Object Marking and Language Change in Heritage Languages*. Cambridge: Cambridge University Press.
- Montrul, S. (2023). Heritage languages: Language acquired, language lost, language regained. *Annual Review of Linguistics*, 9(1), 399-418. <https://doi.org/10.1146/annurev-linguistics-030521-050236>
- Montrul, S., & Foote, R. (2014). Age of acquisition interactions in bilingual lexical access: A study of the weaker language of L2 learners and heritage speakers. *International Journal of Bilingualism*, 18(3), 274-303. <https://doi.org/10.1177/1367006912443431>
- Montrul, S., & Polinsky, M. (2011). Why not HL Speakers? *Linguistic Approaches to Bilingualism*. Amsterdam: John Benjamins.
- Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English bilingual children. *International Journal of Bilingualism*, 11(3), 301-328.
- Montrul, S., & Sánchez-Walker, N. (2013). Differential Object Marking in Child and Adult Spanish Heritage Speakers. *Language Acquisition*, 20(2), 109-132. <https://doi.org/10.1080/10489223.2013.766741>
- Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114(2), 187-202. <https://doi.org/10.1016/j.jecp.2012.09.002>
- Morgan, G. P., Restrepo, A., & Auza, A. (2013). Comparison of Spanish morphology in monolingual and Spanish-English bilingual children with and without language impairment. *Bilingualism: Language and Cognition*, 16, 578-596.
- Morsi, R. (2009). *Specific language impairment in Egyptian Arabic: A preliminary investigation* [Unpublished thesis, University of Reading].
- Müller, N. (2017). Different sources of delay and acceleration in early child bilingualism. *Zeitschrift Für Sprachwissenschaft*, 36(1), 7-30. <https://doi.org/10.1515/zfs-2017-0002>
- Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4(1), 1-21.
- Munro, M., & Mann, V. (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics*, 26(3), 311-341. <https://doi.org/10.1017/S0142716405050198>
- Munson, B., Kurtz, B., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 48(5), 1033-1047. [https://doi.org/10.1044/1092-4388\(2005\)072](https://doi.org/10.1044/1092-4388(2005)072)
- Nag, S., Snowling, M., & Mirković, J. (2018). The role of language production mechanisms in children's sentence repetition: Evidence from an inflectionally rich language. *Applied Psycholinguistics*, 39(2),

- 303-325. <https://doi.org/10.1017/S0142716417000200>
- Naigles, L. (1990). Children use syntax to learn verb meaning. *Journal of Child Language*, 15, 257-272.
- Nakamura, J. (2015). Nonnative maternal input: Language use and errors in a Thai mother's interactions in Japanese with her child. *Japan Journal of Multilingualism and Multiculturalism*, 21(1), 10-26.
- Nayeb, L., Lagerberg, D., Sarkadi, A., Salameh, E. K., & Eriksson, M. (2021). Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta paediatrica (Oslo, Norway: 1992)*, 110(1), 265-272. <https://doi.org/10.1111/apa.15343>
- Nayeb, L., Wallby, T., Westerlund, M., Salameh, E. V., & Sarkadi, A. (2015). Child Healthcare Nurses Believe That Bilingual Children Show Slower Language Development, Simplify Screening Procedures and Delay Referrals. *Acta Paediatrica*, 104, 198-205.
- Nettelbladt, U. (1992). Language impairment in Swedish children: A multidisciplinary approach. In P. Fletcher & D. Hall (Eds.), *Specific speech and language disorders in children: Correlates, characteristics and outcomes* (pp. 138-151). London, UK: Whurr Publishers.
- Neumann, K., Kauschke, C., Fox-Boyer, A., Lüke, C., Sallat, S., & Kiese-Himmel, C. (2024). Clinical practice guideline: Interventions for Developmental Language Delay and Disorders. *Deutsches Arzteblatt international*, 121(5), 155-162. <https://doi.org/10.3238/arztebl.m2024.0004>
- Newbury, D., Winchester, L., Addis, L., Paracchini, S., Buckingham, L., Clark, A., et al. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *The American Journal of Human Genetics*, 85, 1-9. <https://doi.org/10.1016/j.ajhg.2009.07.004>.
- Newbury, J., Bartoszewicz Poole, A., & Theys, C. (2020). Current practices of New Zealand speech-language pathologists working with multilingual children. *International Journal of Speech-Language Pathology*, 22(5), 571-582. <https://doi.org/10.1080/17549507.2020.1712476>
- Newcomer, P. L., & Hammill, D. D. (2008). *Test of Language Development-Primary* (4th ed.). PRO-ED.
- Newcomer, P. L., & Hammill, D. D. (2019). *Test of Language Development-Primary* (5th ed.). Torrance, CA: Western Psychological Services.
- Newport, E. M. (1990). Maturation constraints on language learning. *Cognitive Science*, 14, 11-28.
- Nickisch, A., & Von Kries, R. (2009). Short-term memory constraints in children with specific language impairment (SLI): Are there differences between receptive and expressive SLI? *Journal of Speech, Language and Hearing Research*, 52, 578-595.
- Nicoladis, E. (2016). Bilingual speakers' cognitive development in childhood. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the Life Span: Factors Moderating Language Proficiency* (pp. 269-284). Berlin: de Gruyter Mouton.
- Nicoladis, E. (2018). Simultaneous child bilingualism. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 81-102). Amsterdam: John Benjamins.
- Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science*, 10(2), 237-254. <https://doi.org/10.1111/j.1467-7687.2007.00582.x>.
- Niedo, J., Abbott, R., & Berninger, V. (2014). Predicting levels of reading and writing achievement in typically developing, English-speaking 2nd and 5th graders. *Learning and Individual Differences*, 32, 54-68. <https://doi.org/10.1016/j.lindif.2014.03.013>
- Nippold, M. A., & Fey, S. H. (1983). Metaphoric understanding in preadolescents having a history of language acquisition difficulties. *Language, Speech, and Hearing Services in Schools*, 14(3), 171-180.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic development in adolescents with a history of language impairments: A follow-up investigation. *American Journal of Speech-Language Pathology*, 18, 241-251.
- Nippold, M. A., & Schwarz, I. E. (2002). Do children recover from specific language impairment? *Advances in Speech-Language Pathology*, 1, 41-49.

- Noonan, N. B., Redmond, S. M., & Archibald, L. M. D. (2014). Contributions of children's linguistic and working memory proficiencies to their judgments of grammaticality. *Journal of Speech, Language, and Hearing Research*, 57, 979–989.
- Norbury, C. F., Bishop, D. V. M., & Briscoe, J. (2001). Production of English finite verb morphology: A comparison of SLI and mild-moderate hearing impairment. *Journal of Speech, Language, and Hearing Research*, 44, 165–78.
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of child psychology and psychiatry*, 57(11), 1247-1257. <https://doi.org/10.1111/jcpp.12573>
- Novogrodsky, R. (2015). Specific language impairment (SLI) is not specific enough: sub-types of SLI and their implications for the theory of the disorder. In S. Stavrakaki (Ed.), *Language Acquisition and Language Disorders* (pp.113-124). Baltimore, MD: John Benjamins.
- Novogrodsky, R., & Friedmann, N. (2006). The production of relative clauses in syntactic SLI: A window to the nature of the impairment. *Advances in Speech Language Pathology*, 8(4), 364-375.
- Novogrodsky, R., & Kreiser, V. (2015). What can errors tell us about specific language impairment deficits? Semantic and morphological cuing in a sentence completion task. *Clinical Linguistics & Phonetics*, 29(11), 812–825. <https://doi.org/10.3109/02699206.2015.1051239>.
- Novogrodsky, R., & Meir, N. (2022). Multilingual children with special needs in early education. In M. Schwartz (Ed.), *Handbook of early language education* (pp. 669–696). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-91662-6_18
- Nudel, R., Christiani, C. A., Ohland, J., Uddin, M. J., Hemager, N., Ellersgaard, D. V., ... & Werge, T. (2020). Language deficits in specific language impairment, attention deficit/hyperactivity disorder, and autism spectrum disorder: An analysis of polygenic risk. *Autism Research*, 13(3), 369-381. <https://doi.org/10.1002/aur.2211>.
- Nudel, R., Simpson, N. H., Baird, G., O'Hare, A., Conti-Ramsden, G., Bolton, P. F., ... & Newbury, D. F. (2014). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes, Brain and Behavior*, 13(4), 418-429. <https://doi.org/10.1111/gbb.12127>.
- O'Brien, E. K., Zhang, X., Nishimura, C., Tomblin, J. B., & Murray, J. C. (2003). Association of specific language impairment (SLI) to the region of 7q31. *The American Journal of Human Genetics*, 72(6), 1536-1543.
- Öberg, L., & Bohnacker, U. (2022). Non-Word Repetition and Vocabulary in Arabic-Swedish-Speaking 4–7-Year-Olds with and without Developmental Language Disorder. *Languages*, 7(3), 204. <https://doi.org/10.3390/languages7030204>
- Oetting J. B. (1999). Children with SLI use argument structure cues to learn verbs. *Journal of Speech, Language, and Hearing Research*, 42(5), 1261–1274. <https://doi.org/10.1044/jslhr.4205.1261>
- Oetting, J. B., Cleveland, L. H., & Cope, R. F., 3rd (2008). Empirically derived combinations of tools and clinical cutoffs: an illustrative case with a sample of culturally/linguistically diverse children. *Language, Speech, and Hearing Services in Schools*, 39(1), 44–53. [https://doi.org/10.1044/0161-1461\(2008/005\)](https://doi.org/10.1044/0161-1461(2008/005))
- Oetting, J., Rice, M., & Swank, L. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*, 38, 434–445.
- Öwerdieck, D., Hamann, C., & Abed Ibrahim, L. (2021). In D. Dionne & L. A. Vidal Covas (Eds.), *Proceedings of the 45th annual Boston University Conference on Language Development* (pp. 612-625). Sommerville: Cascadilla Press.
- O'Grady, W. (1997). *Syntactic development*. Chicago, Ill.: University of Chicago Press.
- Okanda M., Moriguchi Y., & Itakura S. (2010). Language and cognitive shifting: evidence from young monolingual and bilingual children. *Psychological Report*. 107, 68–78.
- Oller, D. K., & Eilers, R. E. (2002). *Language and Literacy in Bilingual Children*. London: Multilingual

Matters.

- Oller, D. K., Jarmulowicz, L., Pearson, B., & Cobo-Lewis, A. (2011). Rapid spoken language shift in early second-language learning: The role of peers and effects on the first language. In A. Durgunoglu & C. Goldenberg (Eds.), *Language and literacy development in bilingual settings* (pp. 94–120). New York, NY: Guilford Press
- Oller, K., Pearson, B., & Cobo-Lewis, A. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191–230. <https://doi.org/10.1017/S0142716407070117>
- Omar, M. (1973). *The acquisition of Egyptian Arabic as a native language* (*Janua Linguarum. Series Practica*). Mouton. <https://doi.org/10.1515/9783110819335>
- Orgassa, A. (2009). *Specific language impairment in a bilingual context: The acquisition of Dutch inflection by Turkish-Dutch learners* (Unpublished doctoral dissertation, University of Amsterdam).
- Orgassa, A., & Weerman, F. (2008). Dutch gender in specific language impairment and second language acquisition. *Second Language Research*, 24, 333–364.
- Ortiz, J. A. (2021). Using nonword repetition to identify language impairment in bilingual children: A meta-analysis of diagnostic accuracy. *American Journal of Speech-Language Pathology*, 30(5), 2275–2295. https://doi.org/10.1044/2021_AJSLP-20-00237
- Orsolini, M., Sechi, E., Maronato, C., Bonvino, E., & Corcelli, A. (2001). Nature of phonological delay in children with specific language impairment. *International Journal of Language & Communication Disorders*, 36(1), 63–90. <https://doi.org/10.1080/13682820119370>
- Otheguy, R., & Zentella, A. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.
- Ott, S., van de Vijver, R., & Höhle, B. (2006). The effect of phonotactic constraints in German-speaking children with delayed phonological acquisition. *Advances in Speech and Language Pathology*, 8(4), 323–334. <https://doi.org/10.1080/14417040600970622>
- Ouali, H. (2018). The syntax of tense in Arabic. In E. Benmamoun & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics* (pp. 89–103). Routledge. <https://doi.org/10.4324/9781315147062-6>
- Owen, A. J., & Leonard, L. B. (2006). The production of finite and nonfinite complement clauses by children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research*, 49, 548–571. [https://doi.org/10.1044/10902-4388\(2006/040\)](https://doi.org/10.1044/10902-4388(2006/040))
- Paap, K.R., & Greenberg, Z.I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258. <https://doi.org/10.1016/j.cogpsych.2012.12.002>
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278. <https://doi.org/10.1016/j.cortex.2015.04.014>
- Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in psychology*, 5, 962. <https://doi.org/10.3389/fpsyg.2014.00962>
- Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism*, 5(1), 19–38. <https://doi.org/10.1177/13670069010050010>
- Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 387–406). Blackwell.
- Paradis, J. (2010a). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics*, 31(2), 227–252. <https://doi.org/10.1017/S0142716409990373>
- Paradis, J. (2010b). The interface between bilingual development and specific language impairment: Response to commentaries. *Applied Psycholinguistics*, 31(2), 119–136. <https://doi.org/10.1017/S0142716409990373>
- Paradis, J. (2010c). Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning*, 60(3), 651–680. <https://doi.org/10.1111/j.1467-9922.2010.00567.x>

- Paradis, J. (2011). Individual differences in child English second-language acquisition: Comparing child internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1(3), 213-237. <https://doi.org/10.1075/lab.1.3.01par>
- Paradis, J. (2016). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research*, 59(1), 171–182. https://doi.org/10.1044/2015_JSLHR-L-15-0008
- Paradis, J. (2019). English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In M. Brown & B. Dailey (Eds.), *Proceedings of BUCLD 43* (pp. 11-26). Somerville, MA: Cascadilla Press.
- Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, 50(4), 793–817. <https://doi.org/10.1017/S0305000922000708>
- Paradis, J., & Crago, M. (2001). The Morphosyntax of Specific Language Impairment in French: An Extended Optional Default Account. *Language Acquisition*, 9(4), 269–300. https://doi.org/10.1207/S15327817LA0904_01
- Paradis, J., & Crago, M. (2004). Comparing L2 and SLI grammars in French: Focus on DP. In P. Prevost & J. Paradis (Eds.), *The acquisition of French in different contexts: Focus on functional categories* (pp. 89–108). Amsterdam: John Benjamins.
- Paradis, J., Crago, M., & Genesee, F. (2003). Object clitics as clinical markers of SLI in French: evidence from French-English bilingual children. In B. Beachley, A. Brown & F., *Proceedings of BUCLD 27* (pp. 638-649). Somerville: Cascadilla Press.
- Paradis, J., Crago, M., & Genesee, F. (2005-2006). Domain-General Versus Domain-Specific Accounts of Specific Language Impairment: Evidence From Bilingual Children's Acquisition of Object Pronouns. *Language Acquisition: A Journal of Developmental Linguistics*, 13(1), 33–62. https://doi.org/10.1207/s15327817la1301_3
- Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: using parent report on first language development. *Journal of Communication Disorders*, 43(6), 474–497. <https://doi.org/10.1016/j.jcomdis.2010.01.002>
- Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: autonomous or interdependent? *Studies in Second Language Acquisition*, 18(1), 1–25. <https://doi.org/10.1017/S0272263100014662>
- Paradis, J., Genesee, F., & Crago, M.B. (2021a). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*, 3rd Ed. Baltimore, MD: Brookes.
- Paradis, J., & Grüter, T. (2014). Introduction. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 1–14). Amsterdam: John Benjamins.
- Paradis, J., & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, 1-15. <https://doi.org/10.1111/desc.12433>
- Paradis, J., Jia, R., & Arppe, A. (2017a). The acquisition of tense morphology over time by English second language children with specific language impairment: Testing the cumulative effects hypothesis. *Applied Psycholinguistics*, 38(4), 881-908. <https://doi.org/10.1017/S0142716416000485>
- Paradis, J., & Kirova, A. (2014). English second-language learners in preschool: Profile effects in their English abilities and the role of home language environment. *International Journal of Behavioral Development*, 38(4), 342-349. <http://dx.doi.org/10.1177/0165025414530630>
- Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: what is the role of the input? *Journal of Child Language*, 30(2), 371–393. <https://doi.org/10.1017/S0305000903005609>
- Paradis, J., Nicoladis, E., Crago, M.B., & Genesee, F. (2011). Bilingual children's acquisition of the past tense: a usage-based approach. *Journal of Child Language*, 38(3), 554–578. <https://doi.org/10.1017/S0305000910000218>
- Paradis, J., Rice, M. L., Crago, M.B., & Marquis, J. (2008). The acquisition of tense in English:

- Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, 29(4), 689–722. <https://doi.org/10.1017/S0142716408080296>
- Paradis, J., Rusk, B., Sorenson Duncan, T., & Govindarajan, K. (2017b). Children's second language acquisition of English complex syntax: The role of age, input and cognitive factors. *Annual Review of Applied Linguistics*, 37, 1-20.
- Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, 41(6), 1255-1281.
- Paradis, J., Soto-Corominas, A., Daskalaki, E., Chen, X., & Gottardo, A. (2021b). Morphosyntactic Development in First Generation Arabic—English Children: The Effect of Cognitive, Age, and Input Factors over Time and across Languages. *Languages*, 6(1), 51. <https://doi.org/10.3390/languages6010051>
- Paradis, J., Tremblay, A., & Crago, M.B. (2014). French-English bilingual children's sensitivity to child-level and language-level input factors in morphosyntactic acquisition. *Input and experience in bilingual development*, 13, 161-180.
- Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: individual variation in long-term outcomes. *Journal of Child Language*, 43(3), 553–580. <https://doi.org/10.1017/S0305000915000562>
- Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In: B. Köpke, M. Schmid, M. Keijzer, & S. Dostert (Eds.) *Language attrition: Theoretical perspectives* (pp. 121–34). Amsterdam: John Benjamins.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.
- Patel, A., & Singh, P. (2013). New Approach for K-mean and K-medoids Algorithm. *International Journal of Computer Applications Technology and Research*, 2(1), 1-5. <https://doi.org/10.7753/IJCATR0201.1001>
- Patterson, J. L., & Pearson, B. Z. (2004). Bilingual Lexical Development: Influences, Contexts, and Processes. In B. A. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (pp. 77–104). Paul H. Brookes Publishing Co.
- Paul, R., & Cohen, D. J. (1984). Outcomes of severe disorders of language acquisition. *Journal of Autism and Development Disorders*, 14, 405- 421.
- Paul, R., Cohen, D., & Caparulo, B. (1983). A longitudinal study of patients with severe, specific developmental language disorders. *Journal of the American Academy of Child Psychiatry*, 22, 525-534.
- Pauls, L., & Archibald, L. (2016). Executive Functions in children with specific language impairment: A meta-analysis. *Journal of Speech, Language and Hearing Research*, 59(5), 1074–1086. 1086. https://doi.org/10.1044/2016_JSLHR-L-15-0174
- Pearson, B. Z. (1998). Assessing lexical development in bilingual babies and toddlers. *International Journal of Bilingualism*, 2, 347–372.
- Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied psycholinguistics*, 18(1), 41-58. <https://doi.org/10.1017/S0142716400009863>
- Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43, 93–120. <http://dx.doi.org/10.1111/j.1467-1770.1993.tb00174.x>
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373-1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)
- Peets, K. F., & Bialystok, E. (2010). An integrated approach to the study of specific language impairment and bilingualism. *Applied Psycholinguistics*, 31(2), 315-319. <https://doi.org/10.1017/S0142716409990488>

- Peña E. D., & Bedore L. M. (2008). Bilingualism and Language Impairment in Children. In R. Schwartz (Ed.), *Handbook of Child Language Disorders* (pp.281-307). New York: Psychology Press.
- Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51(2), 192-202.
- Peña, E. D., Bedore, L. M., Shivabasappa, P., & Niu, L. (2020). Effects of divided input on bilingual children with language impairment. *International Journal of Bilingualism*, 24(1), 62-78. <https://doi.org/10.1177/1367006918768367>
- Peña, E., Gillam, R., Bedore, L., & Bohman, T.M. (2011). Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. *American Journal of Speech-Language Pathology*, 20(4), 302–314. [https://doi.org/10.1044/1058-0360\(2011/10-0020\)](https://doi.org/10.1044/1058-0360(2011/10-0020))
- Penner, Z. (1990). On the acquisition of verb-placement and verb projection raising in Bernese Swiss German. In M. Rothweiler (Ed.), *Spracherwerb und Grammatik. Linguistische Untersuchungen zum Erwerb von Syntax und Morphologie [Linguistische Berichte Sonderheft 3]* (pp. 166-189). Opladen: Westdeutscher Verlag.
- Penner, Z. (1994). Asking questions without CPs? On the acquisition of wh questions in Bernese Swiss German and Standard German. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in Generative Grammar* (pp. 177—214). Amsterdam: John Benjamins.
- Penner, Z. (1995a). *Störungen im Erwerb der Nominalphrase im Schweizerdeutschen. Ein spontansprachliches Diagnoseverfahren*. Luzern: Edition SZH.
- Penner, Z. (1995b). *Continuity Constraints on Paradigm Formation in the Acquisition of the Complementizer System. A Case Study in Bernese Swiss German*. Bern: University of Bern
- Penner, Z., & Kölliker Funk, M. (1998). *Therapie und Diagnose von Grammatikerwerbsstörungen. Ein Arbeitsbuch*. Luzern: Ed. SZH/SPC.
- Penner, Z., Kölliker, F., & Zimmermann, H. (1992). *Gestörter Grammatikerwerb im Schweizerdeutschen. [Impaired Grammar Acquisition in Swiss German]*. Luzern: Edition SZH.
- Penner, Z., & Roeper, T. (1998). Trigger theory and the acquisition of complement idioms. In N. Dittmar & Z. Penner (Eds.), *Issues in the Theory of Language Acquisition. Essays in Honor of Jürgen Weissenborn* (pp. 77-112). Bern: Peter Lang.
- Penner, Z., Schönenberger, M., & Weissenborn, J. (1994). The acquisition of object placement in early German and Swiss German. *Linguistics in Potsdam*, 1, 93-108.
- Penner, Z., Schulz, P., & Wymann, K. (2003). Learning the meaning of verbs: what distinguishes language impaired from normally developing children? *Linguistics*, 41(2), 289–319. <https://doi.org/10.1515/ling.2003.010>
- Pérez-Leroux, A., Roberge, Y., & Pirvulescu, M. (2009). Bilingualism as a window into the language faculty: The acquisition of objects in French-speaking children in bilingual and monolingual contexts. *Bilingualism: Language and Cognition*, 12(1), 97–112.
- Pesetsky, David (1987). Wh-in-situ: movement and unselective binding. In E. Reuland, & A. ter Meulen (Eds.), *The Representation of in Definiteness* (pp. 98-129). MIT Press, Cambridge, MA.
- Pesetsky, D., & Torrego, E. (2007). The syntax of valuation and the interpretability of features. In S. Karimi, V. Samian & W. Wilkins (Eds.), *Phrasal and Clausal Architecture: Syntactic derivation and interpretation* (pp. 262-294). Amsterdam: Benjamins.
- Petermann, F., & Petermann, U. (2011). *Wechsler Intelligence Scale for Children* (4th ed.). Frankfurt a. M.: Pearson Assessment.
- Petersen, J. (1988). Word-internal code-switching constraints in a bilingual child's grammar. *Linguistics* 26(3), 479–493. <https://doi.org/10.1515/ling.1988.26.3.479>
- Phillips, C. (1995). Syntax at age two: Cross-linguistic differences. In C. T. Schuetze, J. B. Ganger & K. Broihier (Eds.), *Papers on language processing and acquisition. MIT Working Papers in Linguistics* 26 (pp. 325-382). Cambridge, Mass.: MIT, MITWPL

- Phillips, C. (2010). Syntax at Age Two: Cross-Linguistic Differences. *Language Acquisition*, 17(1–2), 70–120. <https://doi.org/10.1080/10489221003621167>
- Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic Parsing. In Matthew J. Traxler & Morton A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 455–503). Elsevier. <https://doi.org/10.1016/B978-012369374-7/50013-4>
- Pierce, L. J., Genesee, F., Delcenserie, A., & Morgan, G. (2017a). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics*, 38, 1265–1300. <https://doi.org/10.1017/S0142716417000236>
- Pierce, L., Genesee, F., Delcenserie, A., & Morgan, G. (2017b). Toward a model of multiple paths to language learning: Response to commentaries. *Applied Psycholinguistics*, 38(6), 1351–1362. <https://doi.org/10.1017/S0142716417000340>
- Pinker, S. (1991). Rules of Language. *Science*, 253, 530–535.
- Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, 92, 377–410.
- Pinker, S. (1999). *Words and rules: the ingredients of language*. London: Weidenfeld & Nicholson.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense debate. *Trends in Cognitive Sciences*, 6(11), 456–463.
- Pirvulescu, M. (2006). Theoretical implications of object clitic omission in early French: spontaneous vs. elicited production. *Catalan Journal of Linguistics*, 5, 221–236.
- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, 19(5), 1023–1041. <https://doi.org/10.1017/S1366728915000322>
- Plante, E., & Vance, R. (1994). Diagnostic accuracy of two tests of preschool language. *American Journal of Speech-Language Pathology*, 4, 70–76. <https://doi.org/10.1044/1058-0360.0402.70>
- Platzack, C. (1990). A Grammar without Functional Categories: A Syntactic Study of Early Swedish Child Language. *Nordic Journal of Linguistics*, 13(2), 107–126. <https://doi.org/10.1017/S0332586500002183>
- Platzack, C. (1992). Functional categories in early Swedish. In J. Meisel (Ed.), *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition* (pp. 63–82). Dordrecht: Kluwer Academic Publishers.
- Plunkett, K. (1999). Targeting complex structure in French questions. In A. Greenhill, M. Hughes, H. Littlefield & C. Tano (Eds.), *Proceedings of BUCLD 23* (pp. 764–775). Somerville: Cascadilla Press.
- Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 113(4), 535–551. <https://doi.org/10.1016/j.jecp.2012.06.013>
- Poepfel, D., & Wexler, K. (1993). The Full Competence Hypothesis of Clause Structure in Early German. *Language*, 69(1), 1–33. <https://doi.org/10.2307/416414>
- Polinsky, M. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, 14, 191–262.
- Polinsky, M. (2018). Bilingual children and adult heritage speakers: The range of comparison. *International Journal of Bilingualism*, 22(5), 547–563. <https://doi.org/10.1177/136700691665604>
- Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, 23(1), 4–20. <https://doi.org/10.1017/S1366728919000245>
- Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: what does the task measure?. *International Journal of Language & Communication Disorders*, 50(1), 106–118. <https://doi.org/10.1111/1460-6984.12126>
- Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research*, 53(2), 414–429. [https://doi.org/10.1044/1092-4388\(2009/08-0016\)](https://doi.org/10.1044/1092-4388(2009/08-0016))

- Pollock, J. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365–424. <https://www.jstor.org/stable/4178634>
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42. <https://doi.org/10.1146/annurev.ne.13.030190.000325>
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29(6), 633–654. [https://doi.org/10.1016/0749-596X\(90\)90042-X](https://doi.org/10.1016/0749-596X(90)90042-X)
- Poulain, T., & Brauer, J. (2018). The changing role of mothers' verbal and nonverbal behavior in children's language acquisition. *First Language*, 38(2), 129–146. <https://doi.org/10.1177/014272371773542>
- Powell, J., & Wagner, S. (2014). An der Schnittstelle Ethnie und Behinderung benachteiligt. In G. Wansing & M. Westphal (Eds.), *Behinderung und Migration* (pp. 177–199). Wiesbaden: Springer.
- Prevoo, M. J. L., Malda, M., Mesman, J., Emmen, R. A. G., Yeniad, N., Van Ijzendoorn, M. H., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: different pathways for host and ethnic language. *Journal of Child Language*, 41(5), 963–984. <https://doi.org/10.1017/S0305000913000299>
- Prévost, P. (2003). Truncation and missing inflection in initial child L2 German. *Studies in Second Language Acquisition*, 25(1), 65–97.
- Prévost, P., Morin, E., & Tuller, L. (2011). The effect of computational complexity on the production of errors in wh-questions used by children with SLI and L2 children learning French. *Paper presented at IASCL*, Montreal, July 2011.
- Prévost, P., Strik, N., & Tuller, L. (2014). Wh-questions in child L2 French: Derivational complexity and its interactions with L1 properties, length of exposure, age of exposure, and the input. *Second Language Research*, 30(2), 225–250. <https://doi.org/10.1177/0267658313519814>
- Prévost, P., & White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16(2), 103–133. <https://doi.org/10.1191/0267658006775560>
- Prigent, G., Parisse, C., Leclercq, A. L., & Maillart, C. (2015). Complexity markers in morphosyntactic productions in French-speaking children with specific language impairment (SLI). *Clinical Linguistics & Phonetics*, 29(8–10), 701–718. <https://doi.org/10.3109/02699206.2015.1020451>
- Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.
- Pronina, M., & Prieto, P. (2024). Acquisition of Suprasegmental Phonology in Child Bilingualism. In M. Amengual (Ed.), *The Cambridge Handbook of Bilingual Phonetics and Phonology* (pp. 359–382). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009105767.016>
- Qu, L., Low, J. J. W., Zhang, T., Li, H., & Zelazo, P. D. (2016). Bilingual advantage in executive control when task demands are considered. *Bilingualism: Language and Cognition*, 19(2), 277–293. <https://doi.org/10.1017/S1366728914000376>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman.
- Quiroz, B. G., Snow, C. E., & Zhao, J. (2010). Vocabulary skills of Spanish—English bilinguals: Impact of mother—child language interactions and home language and literacy support. *International Journal of Bilingualism*, 14(4), 379–399. <https://doi.org/10.1177/1367006910370919>
- R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Radford, A. (1990). *Syntactic theory and the acquisition of English syntax: The nature of early child grammar of English*. Oxford: Blackwell.
- Rainey, C. (2016). Dealing with Separation in Logistic Regression Models. *Political Analysis*, 24(3), 339–355. <https://doi.org/10.1093/pan/mpw014>

- Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied neuropsychology. Child*, 10(1), 37–52. <https://doi.org/10.1080/21622965.2019.1596113>
- Rakhlin, N., Kornilov, S. A., Kornilova, T. V., & Grigorenko, E. L. (2016). Syntactic Complexity Effects of Russian Relative Clause Sentences in Children with and without Developmental Language Disorder. *Language acquisition*, 23(4), 333–360. <https://doi.org/10.1080/10489223.2016.1179312>
- Rakhlin, N. V., Li, N., Aljughaiman, A., & Grigorenko, E. L. (2020). Narrative Language Markers of Arabic Language Development and Impairment. *Journal of Speech, Language, and Hearing Research*, 63(10), 3472–3487. https://doi.org/10.1044/2020_JSLHR-20-00082
- Rankin, T., & Unsworth, S. (2016). Beyond poverty: Engaging with input in generative SLA. *Second Language Research*, 32(4), 563–572. <https://doi.org/10.1177/02676583166487>
- Rasetti, L. (1996). Null subjects and root infinitives in the child grammar of French. *Geneva Generative Papers*, 4(2), 120–132.
- Rasetti, L. (2000). Null subjects and root infinitives in the child grammar of French. In M.-A. Friedemann & L. Rizzi (Eds.), *The acquisition of syntax* (pp. 236–268). London: Longman.
- Rauch, D. P., Naumann, J., & Jude, N. (2012). Metalinguistic awareness mediates effects of full biliteracy on third-language reading proficiency in Turkish–German bilinguals. *International Journal of Bilingualism*, 16(4), 402–418. <https://doi.org/10.1177/1367006911425>
- Raven, J. C., & Court, J. H. (1986). *Raven's progressive matrices and Raven's coloured matrices*. London: HK Lewis.
- Redmond, S. M., Ash, A. C., Christopoulos, T. T., & Pfaff, T. (2019). Diagnostic accuracy of sentence recall and past tense measures for identifying children's language impairments. *Journal of Speech, Language, and Hearing Research*, 62(7), 2438–2454. https://doi.org/10.1044/2019_JSLHR-L-18-0388
- Rehbein, J., Herkenrath, A., & Karakoç, B. (2009). Turkish in Germany - On contact-induced language change of an immigrant language in the multilingual landscape of Europe. *STUF - Language Typology and Universals*, 62(3), 171–204. <https://doi.org/10.1524/stuf.2009.0011>
- Reilly, S., Bishop, D. V., & Tomblin, J. B. (2014a). Terminological debate over language impairment in children: forward movement and sticking points. *International Journal of Language & Communication Disorders*, 49(4), 452–462. <https://doi.org/10.1111/1460-6984.12111>
- Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and language*, 88(2), 229–247. [https://doi.org/10.1016/S0093-934X\(03\)00101-9](https://doi.org/10.1016/S0093-934X(03)00101-9)
- Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014b). Specific language impairment: a convenient label for whom?. *International Journal of Language & Communication Disorders*, 49(4), 416–451. <https://doi.org/10.1111/1460-6984.12102>
- Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P., & Bretherton, L. (2010). Predicting language outcomes at 4 years of age: findings from Early Language in Victoria Study. *Pediatrics*, 126(6), e1530–e1537. <https://doi.org/10.1542/peds.2010-0254>
- Repovš, G., & Baddeley, A. D. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5–22. <http://dx.doi.org/10.1016/j.neuroscience.2005.12.061>
- Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research*, 39(1), 153–165. <https://doi.org/10.1044/jshr.3901.153>
- Restrepo M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, 41(6), 1398–1411. <https://doi.org/10.1044/jslhr.4106.1398>
- Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry*, 32, 439–492.
- Ribot, K. M., & Hoff, E. (2014). “¿Cómo estas?” “I’m good.” Conversational code-switching is related to

- profiles of expressive and receptive proficiency in Spanish-English bilingual toddlers. *International Journal of Behavioral Development*, 38(4), 333–341. <https://doi.org/10.1177/0165025414533225>
- Ribot, K. M., Hoff, E., & Burrige, A. (2018). Language Use Contributes to Expressive Language Growth: Evidence From Bilingual Children. *Child Development*, 89(3), 929–940. <https://doi.org/10.1111/cdev.12770>
- Rice, M. L. (1991). Children with specific Language impairment: Toward a model of teachability. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 447-480). Hillsdale, NJ: Erlbaum.
- Rice, M. L. (2004). Growth models of developmental language disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207–240). Mahwah, NJ: Erlbaum.
- Rice, M. L. (2020a). Clinical lessons from studies of children with specific language impairment. *Perspectives of the ASHA Special Interest Groups*, 5(1), 12-29. https://doi.org/10.1044/2019_PERSP-19-000
- Rice, M. L. (2020b). Causal pathways for specific language impairment: Lessons from studies of twins. *Journal of Speech, Language, and Hearing Research*, 63(10), 3224-3235. https://doi.org/10.1044/2020_JSLHR-20-00169
- Rice, M. L., Buhr, J., & Oetting, J. B. (1992). Specific-language-impaired children's quick incidental learning of words: the effect of a pause. *Journal of Speech and Hearing Research*, 35(5), 1040–1048. <https://doi.org/10.1044/jshr.3505.1040>
- Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, 58(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150
- Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: a study of growth and asymptote. *Journal of Speech, Language, and Hearing Research*, 52(6), 1417–1433. [https://doi.org/10.1044/1092-4388\(2009/08-0171\)](https://doi.org/10.1044/1092-4388(2009/08-0171))
- Rice, M. L., Noll, K. R., & Grimm, H. (1997). An Extended Optional Infinitive Stage in German-Speaking Children with Specific Language Impairment. *Language Acquisition*, 6(4), 255–295. https://doi.org/10.1207/s15327817la0604_1
- Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, 37(1), 106–122. <https://doi.org/10.1044/jshr.3701.106>
- Rice, M. L., Taylor, C. L., Zubrick, S. R., Hoffman, L., & Earnest, K. K. (2020). Heritability of specific language impairment and nonspecific language impairment at ages 4 and 6 years across phenotypes of speech, language, and nonverbal cognition. *Journal of Speech, Language, and Hearing Research*, 63(3), 793-813. https://doi.org/10.1044/2019_JSLHR-19-00012
- Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research*, 47(4), 816–834. [https://doi.org/10.1044/1092-4388\(2004/061\)](https://doi.org/10.1044/1092-4388(2004/061))
- Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, 39(6), 1239–1257. <https://doi.org/10.1044/jshr.3906.1239>
- Rice, M. L., & Wexler, K. (2001). *Test of Early Grammatical Impairment (TEGI)*. San Antonio, TX: The Psychological Corporation.
- Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*, 38(4), 850–863. <https://doi.org/10.1044/jshr.3804.850>
- Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing*

- Research*, 41(6), 1412–1431. <https://doi.org/10.1044/jslhr.4106.1412>
- Richardson J. T. (2007). Measures of short-term memory: a historical review. *Cortex* 43(5), 635–650. [https://doi.org/10.1016/s0010-9452\(08\)70493-3](https://doi.org/10.1016/s0010-9452(08)70493-3)
- Riches, N. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders*, 47(5), 499–510. <https://doi.org/10.1111/j.1460-6984.2012.00158.x>
- Riches, N. (2020). Commentary on special issue: Syntax and verbal short term memory. *First Language*, 40(4), 500–506. <https://doi.org/10.1177/01427237209352>
- Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: an investigation of complex syntax. *International Journal of Language & Communication Disorders*, 45(1), 47–60. <https://doi.org/10.3109/13682820802647676>
- Rinke, E., & Flores, C. (2021). Portuguese as Heritage Language in Germany—A Linguistic Perspective. *Languages*, 6(1), 10. <https://doi.org/10.3390/languages6010010>
- Rinker, T., Alku, P., Brosch, S., & Kiefer, M. (2010). Discrimination of native and non-native vowel contrasts in bilingual Turkish-German and monolingual German children: Insight from the Mismatch Negativity ERP component. *Brain and Language*, 113(2), 90–95. <https://doi.org/10.1016/j.bandl.2010.01.007>
- Rinker, T., Shafer, V. L., Kiefer, M., Vidal, N., & Yu, Y. H. (2017). T-complex measures in bilingual Spanish-English and Turkish-German children and monolingual peers. *PloS One*, 12(3), e0171992. <https://doi.org/10.1371/journal.pone.0171992>
- Rispens, J., & Baker, A. (2012). Nonword repetition: the relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, 55(3), 683–694. [https://doi.org/10.1044/1092-4388\(2011/10-0263\)](https://doi.org/10.1044/1092-4388(2011/10-0263))
- Rispens, J., Baker, A., & Duinmeijer, I. (2015). Word recognition and nonword repetition in children with language disorders: the effects of neighborhood density, lexical frequency, and phonotactic probability. *Journal of Speech, Language, and Hearing Research*, 58(1), 78–92. https://doi.org/10.1044/2014_JSLHR-L-12-0393
- Rispens, J. E., & De Bree, E. H. (2014). Past tense productivity in Dutch children with and without SLI: the role of morphophonology and frequency. *Journal of Child Language*, 41(1), 200–225. <https://doi.org/10.1017/S0305000912000542>
- Rispens, J., & de Bree, E. (2015). Bilingual children's production of regular and irregular past tense morphology. *Bilingualism: Language and Cognition*, 18(2), 290–303. <https://doi.org/10.1017/S1366728914000108>
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press
- Rizzi, L. (1994). Some notes on linguistic theory and language development: The case of root infinitives. Longman. *Language Acquisition*, 3(4), 371–393.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Dordrecht: Kluwer Academic Publishers.
- Rizzi, L. (2000). Remarks on Early Null Subjects. In M.A. Friedeman, & L. Rizzi (Eds.), *The acquisition of syntax* (pp. 269–292). London: Addison-Wesley
- Rizzi, L. (2004). Locality and left periphery. In A. Belletti (Ed.), *Structures and Beyond. The Cartography of Syntactic Structures* (pp. 223–251), Volume 3. Oxford: OUP.
- Rizzi, L. (2013). Locality. *Lingua*, 130, 169–186. <https://doi.org/10.1016/j.lingua.2012.12.002>
- Rizzi, L. (2018). Intervention effects in grammar and language acquisition. *Probus*, 30(2), 339–367. <https://doi.org/10.1515/probus-2018-0006>
- Roberts, J., Rescorla, L., Giroux, J., & Stevens, L. (1998). Phonological skills of children with specific expressive language impairment (SLI-E): outcome at age 3. *Journal of Speech, Language, and Hearing*

- Research*, 41(2), 374–384. <https://doi.org/10.1044/jslhr.4102.374>
- Rodina, Y., & Westergaard, M. (2015). Grammatical gender in Norwegian: language acquisition and language change. *Journal of Germanic Linguistics*, 27(2), 145–187. <https://doi.org/10.1017/S1470542714000245>
- Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian–Russian acquisition: The role of input and transparency. *Bilingualism: Language and Cognition*, 20(1), 197–214. <https://doi.org/10.1017/S1366728915000668>
- Roeper, T. (1996). The role of merger theory and formal features in acquisition. In H. Clahsen (Ed.), *Generative perspectives on language acquisition [Language Acquisition and Language Disorders 14]* (pp. 415–450). Amsterdam: John Benjamins.
- Roeper, T. (2004). Diagnosing language variations: underlying principles for syntactic assessment. *Seminars in speech and language*, 25(1), 41–55. <https://doi.org/10.1055/s-2004-824825>
- Roeper, T. (2011). The acquisition of recursion: how formalism articulates in the child's path. *Biolinguistics*, 5(1-2), 57–86. <https://doi.org/10.5964/bioling.8831>
- Roeper, T., Ramos, E., Seymour, H., & Abdul-Karim, L. (2001). Language disorders as a window on universal grammar: an abstract theory of agreement for IP, DP, and V-PP. *Brain and language*, 77(3), 378–397. <https://doi.org/10.1006/brln.2000.2409>
- Roesch, A. D., & Chondrogianni, V. (2015). The use of case in the comprehension of wh-questions in German-speaking children with and without SLI. In C. Hamann & E. Ruigendijk (Eds.), *Proceedings of GALA 2013* (pp. 379–402). Cambridge: Cambridge Scholars.
- Roesch, A. D., & Chondrogianni, V. (2016). "Which mouse kissed the frog?" Effects of age of onset, length of exposure, and knowledge of case marking on the comprehension of wh-questions in German-speaking simultaneous and early sequential bilingual children. *Journal of Child Language*, 43(3), 635–661. <https://doi.org/10.1017/S0305000916000015>
- Romonath, R. (1991). *Phonologische Prozesse an sprachauffälligen Kindern: Eine vergleichende Untersuchung an sprachauffälligen und nichtsprachauffälligen Vorschulkindern*. Berlin: Ed. Marhold im Wiss.-Verl. Spiess.
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, 13(2), 155–163. <https://doi.org/10.1177/1367006909339814>
- Rothman, J., Long, D., Iverson, M., Judy, T., Lingwall, A., & Chakravarty, T. (2016). Older age of onset in child L2 acquisition can be facilitative: evidence from the acquisition of English passives by Spanish natives. *Journal of Child Language*, 43(3), 662–686. <https://doi.org/10.1017/S0305000915000549>
- Rothman, J., & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics*, 35(1), 93–98. <https://doi.org/10.1093/applin/amt049>
- Rothweiler, M. (1999). Neue Ergebnisse zum fast mapping bei sprachnormalen und sprachentwicklungsgestörten Kindern. In J. Meibauer & M. Rothweiler (Eds.), *Das Lexikon im Spracherwerb* (pp. 252–277). Tübingen: A. Francke Verlag.
- Rothweiler, M. (2006). The acquisition of V2 and subordinate clauses in early successive acquisition of German. In C. Lle'o (Ed.), *Interfaces in multilingualism. Acquisition and representation. Hamburg Studies on Multilingualism* (pp. 91–113). Amsterdam: John Benjamins. <https://doi.org/10.1075/hsm.4.05rot>
- Rothweiler, M. (2013). Spezifische Sprachentwicklungsstörungen bei mehrsprachigen Kindern. *Sprache-Stimme-Gehör*, 37, 186–190. <http://dx.doi.org/10.1055/s-0033-1358699>
- Rothweiler, M., Chilla, S., & Clahsen, H. (2012). Subject verb agreement in Specific Language Impairment: A study of monolingual and bilingual German-speaking children. *Bilingualism: Language and Cognition*, 15, 39–57.
- Rothweiler, M., & Clahsen, H. (1993). Dissociations in SLI children's inflectional systems: A study of participle inflection and subject-verb-agreement. *Scandinavian Journal of Logopedics and Phoniatrics*,

18(4), 169–179. <https://doi.org/10.3109/14015439309101363>

- Rothweiler, M., Schönenberger, M., & Sterner, F. (2017). Subject-verb agreement in German in bilingual children with and without SLI. *Zeitschrift für Sprachwissenschaft*, 36(1), 79–106. <https://doi.org/10.1515/zfs-2017-0005>
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child directed speech in vocabulary development. *Child Development*, 83(5), 1762–1774. <https://doi.org/10.1111/j.1467-8624.2012.01805.x>
- Rowe, M. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives*, 12, 122–127.
- Rowe, M., & Snow, C. (2020). Analyzing input quality along three dimensions: interactive, linguistic and conceptual. *Journal of Child Language*, 47, 5–21.
- Roy, P., Chiat, S., & Dodd, B. (2014). *Language and socioeconomic disadvantage: From research to practice*. London: City University London.
- Royal College of Speech and Language Therapists Specific Interest Group in Bilingualism. (2007). *Good practice for speech and language therapists working with clients from linguistic minority communities*. London: RCSLT.
- Rubach, J. (1997). Extrasyllabic consonants in Polish: Derivational optimality theory. In I. Roca, *Derivations and constraints in phonology* (pp. 551–581). Cambridge: Cambridge University Press.
- Ruberg, T., Rothweiler, M., Verissimo, J., & Clahsen, H. (2019). Childhood bilingualism and Specific Language Impairment: A study of the CP-domain in German SLI. *Bilingualism: Language and Cognition*, 1–13. <https://doi.org/10.1017/S1366728919000580>
- Ruigendijk, E. (2015). Contrastive elicitation task for testing case marking. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing multilingual children* (pp. 38–54). Bristol: Multilingual Matters. <https://doi.org/10.21832/9781783093137-004>
- Ruigendijk, E., & Friedmann, N. (2017). A deficit in movement-derived sentences in German-speaking hearing-impaired children. *Frontiers in psychology*, 8, 689. <https://doi.org/10.3389/fpsyg.2017.00689>
- Ruiz-Felter, R., Cooperson, S. J., Bedore, L. M., & Peña, E. D. (2016). Influence of current input–output and age of first exposure on phonological acquisition in early bilingual Spanish–English-speaking kindergarteners. *International Journal of Language & Communication Disorders*, 51(4), 368–383. <https://doi.org/10.1111/1460-6984.12214>
- Rujas, I., Mariscal, S., Murillo, E., & Lázaro, M. (2021). Sentence Repetition Tasks to Detect and Prevent Language Difficulties: A Scoping Review. *Children*, 8(7), 578. <https://doi.org/10.3390/children8070578>
- Ryding, K. (2005). *A Reference Grammar of Modern Standard Arabic* (Reference Grammars). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486975.017>
- Saad, S., & Henry, G. (2020). Épreuve Répétitions de phrases: LITMUS-SR-Liban. In R. Zebib, P. Prévost, L. Tuller & G. Henry (Eds.), *Plurilinguisme et Troubles Spécifiques du Langage au Liban [Multilingualism and Specific Language Impairment in Lebanon]* (pp. 57–77). Beirut: Presses universitaires de l'Université Saint Joseph. <https://de.calameo.com/read/0061189916ca7c0fbbc03>
- Saadah, E. (2011). *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic [Unpublished doctoral dissertation]*. University of Illinois, Urbana–Champaign
- Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics*, 24(3), 431–451. <https://doi.org/10.1017/S0142716403000225>
- Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudowords in a diglossic context. *Applied Psycholinguistics*, 25(4), 495–512. <https://doi.org/10.1017/S0142716404001249>
- Saiegh-Haddad, E. (2007). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics*, 28(4), 607–625. <https://doi.org/10.1017/S0142716407070336>
- Saiegh-Haddad E. (2018). MAWRID: A Model of Arabic Word Reading in Development. *Journal of*

- learning disabilities*, 51(5), 454–462. <https://doi.org/10.1177/0022219417720460>
- Saiegh-Haddad, E. (2020). The impact of linguistic distance on dyslexia in dialect speakers: The case of Arabic diglossia. In J. Washington, D. Compton, & P. McCardle (Eds.), *Dyslexia: Revisiting etiology, diagnosis, and treatment* (pp. 90–101). Brookes.
- Saiegh-Haddad, E., & Armon-Lotem, S. (2024). Diglossia and developmental language disorder (DLD) in Arabic: The role of linguistic distance and linguistic proximity? In K. K. Grohmann (Ed.), *Studies in bilingualism* (Vol. 66, pp. 154–183). John Benjamins.
- Saiegh-Haddad, E., & Ghawi-Dakwar, O. (2017). Impact of diglossia on word and non-word repetition among language impaired and typically developing Arabic native speaking children. *Frontiers in psychology*, 8, 2010. <https://doi.org/10.3389/fpsyg.2017.02010>
- Saiegh-Haddad, E., Hadieh, A., & Ravid, D. (2012). Acquiring Noun Plurals in Palestinian Arabic: Morphology, Familiarity, and Pattern Frequency. *Language Learning*, 62, 1079–1109. <https://doi.org/10.1111/j.1467-9922.2012.00727.x>
- Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad & R. Malatesha Joshi (Eds.), *Handbook of Arabic Literacy* (pp. 3–28). Dordrecht: Springer. https://doi.org/10.1007/978-94-017-8545-7_1
- Saiegh-Haddad, E., Shahbari-Kassem, A., & Schiff, R. (2020). Phonological awareness in Arabic: The role of phonological distance, phonological-unit size, and SES. *Reading and Writing*, 33(6), 1649–1674. <https://doi.org/10.1007/s11145-020-10019-3>
- Salameh, E.K., Nettelbladt, U., Hakansson, G., & Gullberg, B. (2002). Language impairment in Swedish bilingual children. A comparison between bilingual and monolingual children. *Acta Paediatrica*, 9(2), 229–234.
- Sauerland, U., Grohmann, K. K., Guasti, M. T., Anđelković, D., Argus, R., Armon-Lotem, S., ... & Yatsushiro, K. (2016). How do 5-year-olds understand questions? Differences in languages across Europe. *First Language*, 36. <http://doi.org/10.1177/0142723716640236>
- Schachar, R., Logan, G. D., Robaey, P., Chen, S., Ickowicz, A., & Barr, C. (2007). Restraint and cancellation: multiple inhibition deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, 35(2), 229–238. <https://doi.org/10.1007/s10802-006-9075-2>
- Scharff, W. (2013). *Kindliche Mehrsprachigkeit. Grundlagen und Praxis der sprachtherapeutischen Intervention*. Stuttgart: Thieme
- Scharff, W. (2019). Speech and Language Therapy Services for Multilingual Children with Migration Background: A Cross-Sectional Survey in Germany. *Folia Phoniatrica et Logopaedica*, 71, 116–126.
- Scarpino, S. E., Hammer, C. S., Goldstein, B., Rodriguez, B. L., & Lopez, L. M. (2019). Effects of Home Language, Oral Language Skills, and Cross-Linguistic Phonological Abilities on Whole-Word Proximity in Spanish-English-Speaking Children. *American Journal of Speech-Language Pathology*, 28(1), 174–187. https://doi.org/10.1044/2018_AJSLP-18-0050
- Scheele, A., Leseman, P., & Mayo, A. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31, 117–140
- Scheidnes, M. (2020). Sentence repetition and non-word repetition in early total French immersion. *Applied Psycholinguistics*, 41(1), 107–131. <https://doi.org/10.1017/S0142716419000420>
- Scheidnes, M., & Tuller, L. (2014). L2 children embed normally. *New directions in the acquisition of Romance languages: Selected proceedings of the Romance Turn V*, 261.
- Scheidnes, M., & Tuller, L. (2019). Using clausal embedding to identify language impairment in sequential bilinguals. *Bilingualism: Language and Cognition*, 22(5), 949–967. <https://doi.org/10.1017/S1366728918000949>
- Scherger, A. L. (2016). Kasuserwerb bilingual deutsch-italienischer Kinder: vorübergehend verzögernder Spracheneinfluss. *Linguistische Berichte*, 246, 60–104. https://doi.org/10.46771/2366077500246_3
- Scherger, A. L. (2018). German dative case marking in monolingual and simultaneous bilingual children with and without SLI. *Journal of Communication Disorders*, 75, 87–101.

<https://doi.org/10.1016/j.jcomdis.2018.06.004>

- Scherger, A.-L. (2019a). Dative Case Marking in 2L1 and L2 Bilingual SLI. In P. Guijarro Fuentes & C. Suárez-Gómez (Eds.), *Language acquisition and development*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing (pp. 95-115).
- Scherger, A. L. (2019b). Elicited Subject-Verb Agreement in German Early L2 Children with Developmental Language Disorders. *Journal of the European Second Language Association*, 3(1), 46-57. <https://doi.org/10.22599/jesla.57>
- Scherger, A. L. (2022). The role of age and timing in bilingual assessment: non-word repetition, subject-verb agreement and case marking in L1 and eL2 children with and without SLI. *Clinical Linguistics & Phonetics*, 36(1), 54–74. <https://doi.org/10.1080/02699206.2021.1885497>
- Schiff, R., & Saiegh-Haddad, E. (2018). Development and Relationships Between Phonological Awareness, Morphological Awareness and Word Reading in Spoken and Standard Arabic. *Frontiers in psychology*, 9, 356. <https://doi.org/10.3389/fpsyg.2018.00356>
- Schipke, C. S., Friederici, A. D., & Oberecker, R. (2011). Brain responses to case-marking violations in German preschool children. *NeuroReport*, 1. <http://doi.org/10.1097/WNR.0b013e32834c1578>
- Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science*, 15(6), 762–774. <https://doi.org/10.1111/j.1467-7687.2012.01167.x>
- Schlyter, S. (1993). The weaker language in bilingual Swedish–French children. In K. Hyltenstam, & A. Viberg (Eds.), *Progression and regression in language: Sociocultural, neuropsychological and linguistic perspectives* (pp. 289-308). Cambridge: CUP.
- Schmid, M., Gilbers, S., & Nota, A. (2014). Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch-English bilingualism. *Second Language Research*, 30(2), 129-157.
- Schmid, M., & Köpke, B. (2013). Second language acquisition and attrition. In M. Schmid & B. Köpke (Eds.), *First language attrition* (pp.1–12). Amsterdam: John Benjamins.
- Schmid, M., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, 7(6), 637–667.
- Schmitz, K., Patuto, M., & Müller, N. (2012). The null-subject parameter at the interface between syntax and pragmatics: Evidence from bilingual German–Italian, German–French, German–French and Italian–French children. *First Language*, 32(1–2), 205–238. <https://doi.org/10.1177/0142723711403880>
- Schneider, P., Dubé, R. V., & Hayward, D. (2006). *Storytelling from pictures using the Edmonton Narrative Norms Instrument*. Canadian Journal of Speech-Language Pathology and Audiology, 30, 224–238.
- Schönenberger, M., Rothweiler, M., & Sterner, F. (2012). Case marking in child L1 and early child L2 German. In K. Braunmüller & C. Gabriel (Eds.), *Multilingual individuals and multilingual societies. Hamburg Studies on Multilingualism 13* (pp. 3–22). Amsterdam: John Benjamins.
- Schouwenaars, A., Hendriks, P., & Ruigendijk, E. (2018). German children's processing of morphosyntactic cues in wh questions. *Applied Psycholinguistics*, 39(6), 1279-1318. <https://doi.org/10.1017/S0142716418000334>
- Schouwenaars, A., van Hout, A., & Hendriks, P. (2014). Word order overrules number agreement: Dutch children's interpretation and production of which-questions. In C. Y. Chu, C. E. Coughlin, B. L. P., U. Minai & A. Tremblay (Eds.), *Selected proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America (GALANA 2012)*, 60–71. Somerville: Cascadilla Proceedings Project.
- Schulz, P. (2007). Erstspracherwerb Deutsch: Sprachliche Fähigkeiten von eins bis zehn. In Graf & E. Moser Opitz (Eds.), *Diagnostik und Förderung im Elementarbereich und Grundschulunterricht: Lernprozesse wahrnehmen, deuten und begleiten* (pp. 67-85). Baltmannsweiler: Schneider Verlag Hohengehren.
- Schulz, P. (2010). Who answered what to whom? On children's understanding of exhaustive questions.

Paper presented at the final conference of COST Action A33, London.

- Schulz, P. (2013a). Differences between typically developing children and children with SLI in a bilingual context: Exhaustivity. *Paper presented at the final conference of COST IS0804*, Krakow, Poland.
- Schulz, P. (2013b). Wer versteht wann was? Sprachverstehen im frühen Zweitspracherwerb des Deutschen am Beispiel der w-Fragen. In Deppermann, A. (Ed.) *Das Deutsch der Migranten* (pp. 313-337). Jahrbuch 2012 des Instituts für Deutsche Sprache. Mannheim: de Gruyter.
- Schulz, P. (2015). 4. Comprehension of Exhaustive Wh-Questions. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 76-94). Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781783093137-006>
- Schulz, P. (2018). Telicity in typical and impaired acquisition. In K. Syrett & S. Arunachalam (Eds.), *Semantics in Language Acquisition* (pp. 123-150). Amsterdam: John Benjamins.
- Schulz, P., & Grimm, A. (2012). Spracherwerb. In H. Drügh & H.-H. Ewers (Eds.), *Germanistik: Sprachwissenschaft – Literaturwissenschaft – Schlüsselkompetenzen* (pp. 155-172). Stuttgart: Metzler. https://doi.org/10.1007/978-3-476-00399-7_6
- Schulz, P., & Grimm, A. (2019). The age factor revisited: Timing in acquisition interacts with age of onset in bilingual children. *Frontiers in Psychology*, 9, 2732. <https://doi.org/10.3389/fpsyg.2018.02732>
- Schulz, P., Grimm, A., Schwarze, R., & Wojtecka, M. (2017). Spracherwerb bei Kindern mit Deutsch als Zweitsprache: Chancen und Herausforderungen. In U. Hartmann, M. Hasselhorn & A. Gold (Eds.), *Entwicklungsverläufe verstehen–Kinder mit Bildungsrisiken wirksam fördern. Forschungsergebnisse des Frankfurter IDeA-Zentrums* (pp. 190- 206). Stuttgart: Kohlhammer.
- Schulz, P., & Kiese-Himmel, C. (2006). Verbverstehen und expressiver Wortschatzumfang bei sprech-/sprachentwicklungsgestörten Kindern. *L.O.G.O.S. Interdisziplinär*, 4, 244-252.
- Schulz, P., & Penner, Z. (2002). How you can eat the apple and have it too: Evidence from the acquisition of telicity in German. In J. Costa & M. J. Freitas (Eds.), *Proceedings of the GALA 2001 Conference on Language Acquisition*, 239-246. Lisbon, Portugal: University of Lisbon.
- Schulz, P., Penner, Z., & Wymann, K. (2002). Comprehension of resultative verbs in normally developing and language impaired children. In F. Windsor, M. L. Kelly, & N. Hewlett (Eds.), *Investigations in clinical phonetics and linguistics* (pp. 115-129). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schulz, P., & Roeper, T. (2011). Acquisition of exhaustivity in wh-questions: a semantic dimension of SLI? *Lingua* 121(3), 383-407. <https://doi.org/10.1016/j.lingua.2010.10.005>
- Schulz, P., & Tracy, R. (2011). *Linguistische Sprachstandserhebung—Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe Verlag.
- Schulz, P., & Schwarze, R. (2017). How strong is the ban on non-finite verbs in V2? Evidence from early second language learners of German with and without SLI. *Zeitschrift Für Sprachwissenschaft*, 36(1), 51-78. <https://doi.org/10.1515/zfs-2017-0004>
- Schulz, P., & Wittek, A. (2003). Opening doors and sweeping floors: What children with specific language impairment know about telic and atelic verbs. In B. Beachley, A. Brown, & F. Colin (Eds.), *Proceedings of the 27th annual Boston University Conference on Language Development* (pp. 727-738). Somerville, MA: Cascadilla Press.
- Schulz, P., Wymann, K., & Penner, Z. (2001). The early acquisition of verb meaning in German by normally developing and language impaired children. *Brain and Language*, 77, 407-418. <https://doi.org/10.1006/brln.2000.2411>
- Schwartz, B. D. (1998). The second language instinct. *Lingua*, 106(1-4), 133-160.
- Schwartz, B. D. (2004). On child L2 development of syntax and morphology. *Lingue e Linguaggio*, 3(1), 97-132. <https://doi.org/10.1418/13315>
- Schwartz, M., & Rovner, H. (2015). The acquisition of definiteness in Hebrew (L2) by bilingual preschool children (Russian-L1): A longitudinal multiple-case study. *International Journal of Bilingualism*, 19(5), 548-571. <https://doi.org/10.1177/1367006914526073>

- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1), 40–72.
- Schwartz, R. G. (2009). Specific language impairment. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 3–43). Psychology Press.
- Schwarze, R., Wojtecka, M., Grimm, A., & Schulz, P. (2015). Finiteness and verb placement in early second language learners with SLI. In C. Hamann & E. Ruigendijk (Eds.), *Language acquisition and development* (pp. 429–445). Cambridge: Cambridge Scholars Publishing.
- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 64(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552
- Scontras, G., Polinsky, M., & Fuchs, Z. (2018). In support of representational economy: Agreement in heritage Spanish. *Glossa: journal of general linguistics*, 3(1). <https://doi.org/10.5334/gjgl.164>
- Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, 43(2), 324–339. <https://doi.org/10.1044/jslhr.4302.324>
- Sekali, M. (2012). The emergence of complex sentences in a French child's language from 0;10 to 4;01: causal adverbial clauses and the concertina effect. *Journal of French Language Studies*, 22(1), 115–141. <https://doi.org/10.1017/S0959269511000615>
- Selkirk, E. O. (1984). *Phonology and syntax. The relation between sound and structure*. Cambridge, MA: MIT Press.
- Semel, E., Wiig, E.H., & Secord, W.A. (1995). *Clinical Evaluations of Language Fundamentals 3 (CELF-3): Examiner's Manual*. San Antonio, TX: Psychological Corporation.
- Semel, E., Wiig, E.H., & Secord, W.A. (2003). *Clinical Evaluations of Language Fundamentals 4 (CELF-4): Examiner's Manual*. San Antonio, TX: Psychological Corporation.
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, 3(1), 3–25. <https://doi.org/10.1075/lab.3.1.01ser>
- Seymour, H., Roeper, T., de Villiers, J., & de Villiers, P. (2005). *Diagnostic Evaluation of Language Variation (DELV™)—Norm-Referenced*. San Antonio: Harcourt Assessment
- Shaanan, S. (2010). *Investigating grammatical complexity in Gulf Arabic-speaking children with specific language impairment (SLI)* [Unpublished doctoral thesis, University College London]. <https://discovery.ucl.ac.uk/id/eprint/20472>
- Shaanan, S. (2017). Reliability and validity of four Arabic language tests: A comparison of performance of Qatari school-aged children with and without language impairment. *Arab Journal of Applied Linguistics*, 2(1), 20–48. <https://arjals.com/ajal/article/view/101>
- Shaanan S. (2020). Nonword Repetition Skills in Gulf Arabic-Speaking Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, 63(11), 3700–3713. https://doi.org/10.1044/2020_JSLHR-20-00040
- Sharwood Smith, M., & van Buren, P. (1991). First language attrition and the parameter setting model. In H. W. Seliger & R. M. Vago (Eds.), *First Language Attrition* (pp. 17–30). Cambridge: Cambridge University Press.
- Sheng, L., & McGregor, K. K. (2010). Lexical–semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 53, 146–159. [https://doi.org/10.1044/1092-4388\(2009/08-0160\)](https://doi.org/10.1044/1092-4388(2009/08-0160))
- Silva-Corvalán, C. (1986). Bilingualism and Language Change: The Extension of Estar in Los Angeles Spanish. *Language*, 62(3), 587–608. <https://doi.org/10.2307/415479>
- Shipstead Z., Harrison T.L., Engle R.W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, 77, 1863–1880.
- Shlonsky, U. (1992). Resumptive pronouns as a last resort. *Linguistic inquiry*, 23(3), 443–468.

- Shlonsky, U. (1997). *Clause Structure and Word Order in Hebrew and Arabic: An Essay in Comparative Semitic Syntax*. Oxford University Press.
- Siegmüller, J. (2014). Entwicklung der Grammatik. In A. Fox-Boyer (Ed.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen- Kindergartenphase* (pp.15-23). München: Elsevier/ Urban & Fischer.
- Simkin, Z., & Conti-Ramsden, G. (2006). Evidence of reading difficulties in subgroups of children with specific language impairment. *Child Language Teaching and Therapy*, 22, 315–331.
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174–176. <https://doi.org/10.1037/h0027448>
- Simos, P., Sideridis, G., Mouzak, A., & Chatzidaki, A. (2014). Vocabulary growth in second language among immigrant school-aged children in Greece. *Applied Psycholinguistics*, 35, 621–647.
- Šimundić A. M. (2009). Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC*, 19(4), 203–211.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Singleton, D. (2001). Age and second language acquisition. *Annual Review of Applied Linguistics*, 21, 77–89. <https://doi.org/10.1017/S0267190501000058>
- Skehan, P. (2012). Language aptitude. In S. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 381-395). London: Routledge.
- SLI Consortium. (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *The American Journal of Human Genetics*, 70(2), 384-398.
- Slobin, D. I. (1985). Crosslinguistic evidence for the Language Making Capacity. In D.I. Slobin (Ed.), *The crosslinguistic study of language acquisition (2nd vol.): Theoretical issues* (pp. 1157–1256). Hillsdale, NJ: Lawrence Erlbaum.
- Smeets, E., Driessen, G., Elfering, S., & Hovius, M. (2009). *Allochtone leerlingen en speciale onderwijsvoorzieningen*. Nijmegen: ITS.
- Smidts, D. P., Jacobs, R., & Anderson, V. (2004). The Object Classification Task for Children (OCTC): A measure of concept generation and mental flexibility in early childhood. *Developmental neuropsychology*, 26(1), 385-401.
- Smith, N., & Tsimpli, I.M. (1995). *The Mind Of A Savant: Language Learning and Modularity*. Oxford, UK and Cambridge, MA: Basil Blackwell.
- Smithson, L., Paradis, J., & Nicoladis, E. (2014). Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference? *Bilingualism: Language and Cognition*, 17(4), 810-821. <https://doi.org/10.1017/S1366728913000813>
- Snowling, M. (2000). *Dyslexia*. Oxford: Blackwell Publishers.
- Snowling, M., Adams, J., Bishop, D. V. M., & Stothard, S. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International Journal of Language & Communication Disorders*, 36(2), 173–183. <https://doi.org/10.1080/13682820010019892>
- Snowling, M., Bishop, D. V. M., Stothard, S., & Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry*, 47, 759–765.
- Snowling M., & Hulme C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society*, 346(1315), 21–27. <https://doi.org/10.1098/rstb.1994.0124>
- Snyder, W. (1995). *Language Acquisition and Language Variation: The Role of Morphology*. PhD dissertation, MIT.
- Soares, C. (2006). *La syntaxe de la périphérie gauche et son acquisition en Portugais Européen* (Unpublished doctoral dissertation). University of Paris 8, Paris.
- Soltan, U. (2010). On strategies of question-formation and the grammatical status of the Q-particle huwwa in Egyptian Arabic wh-questions. *University of Pennsylvania Working Papers in Linguistics*, 17(1), 1–10.

- Song, J. Y., Sundara, M., & Demuth, K. (2009). Phonological constraints on children's production of English third person singular -s. *Journal of Speech, Language and Hearing Research*, 52, 623–642.
- Soni, K. G., & Patel, A. (2017). Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. *International Journal of Computational Intelligence Research*, 13, 899–906.
- Sopata, A. (2011). Placement of infinitives in successive child language acquisition. In E. Rinke & T. Kupisch (Eds.), *The Development of Grammar: Language acquisition and diachronic change. Volume in honour of Jürgen M. Meisel* (pp. 105–121). Amsterdam: John Benjamins. <https://doi.org/10.1075/hsm.11.06sop>
- Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, 1, 1–33.
- Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, 119(3), 460–477. <https://doi.org/10.1016/j.lingua.2008.09.008>
- Sorenson Duncan, T., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language, and Hearing Research*, 59, 39–48. https://doi.org/10.1044/2015_JSLHR-L-14-0020
- Sorenson Duncan, T., & Paradis, J. (2019). Using television to boost native-speaker input for L2-learning children: A cautionary tale. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd Boston University Conference on Language Development* (pp. 639–652). Somerville: Cascadia Press.
- Sorenson Duncan, T., & Paradis, J. (2020a). How does maternal education influence the linguistic environment supporting bilingual language development in child L2 learners of English? *International Journal of Bilingualism*, 24(1), 46–61. <https://doi.org/10.1177/1367006918768366>
- Sorenson Duncan, T., & Paradis, J. (2020b). Home language environment and children's second language acquisition: The special status of input from older siblings. *Journal of Child Language*, 47(5), 982–1005. <https://doi.org/10.1017/S0305000919000977>
- Soto-Corominas, A., Daskalaki, E., Paradis, J., Winters-Difani, M., & Janaideh, R. A. (2022). Sources of variation at the onset of bilingualism: The differential effect of input factors, AOA, and cognitive skills on HL Arabic and L2 English syntax. *Journal of Child Language*, 49(4), 741–773. <https://doi.org/10.1017/S0305000921000246>
- Spaulding, T. J. (2010). Investigating mechanisms of suppression in preschool children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 53, 725–738.
- Spaulding T.J., Plante E., & Vance R. (2008). Sustained selective attention skills of preschool children with specific language impairment: Evidence for separate attentional capacities. *Journal of Speech, Language, and Hearing Research*, 51, 16–34 <https://doi.org/10.1016/j.heares.2013.12.001>
- Stadthagen-González, H., Gathercole, V. C. M., Pérez-Tattam, R., & Yavas, F. (2013). Vocabulary assessment of bilingual adults: To cognate or not to cognate. In V. C. M. Gathercole (Ed.), *Solutions for the assessment of bilinguals* (pp. 125–145). Bristol, England: Multilingual Matters.
- Stanford, E., & Delage, H. (2020). Complex syntax and working memory in children with specific learning difficulties. *First Language*, 40, 411–436.
- Starke, M. (2001). *Move Dissolves into Merge: A Theory of Locality* [Doctoral dissertation, University of Geneva]. <http://www.ai.mit.edu/projects/dm/starke-diss.pdf>
- Statistisches Bundesamt (2019). Bevölkerung: Migration und Integration. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/_inhalt.html;jsessionid=D2B911B5C3637141E5338D5598765F95.live732 (accessed on 22 June 2019).
- Stavarakaki, S. (2001). Comprehension of reversible relative clauses in specifically language impaired and normally developing Greek children. *Brain and Language*, 77(3), 419–431. <https://doi.org/10.1006/brln.2000.2412>
- Stavarakaki, S., Tasioudi, M., & Guasti, T. (2015). Morphological cues in the comprehension of relative clauses by Greek children with specific language impairment and typical development: A comparative

- study. *International Journal of Speech-Language Pathology*, 17(6), 617–626. <https://doi.org/10.3109/17549507.2015.1048826>
- Stavrakaki, S., Vogindroukas, I., Chelas, E., & Ghousi, S. (2008). Subject-verb agreement in SLI and child L2: A comparative approach to Greek SLI and (un) impaired child L2. In A. Gavarró & M. J. Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007* (pp. 413–423). Newcastle upon Tyne: Cambridge Scholars.
- Steenge, J. (2006). *Bilingual children with specific language impairment: Additionally disadvantaged?* Doctoral dissertation, Research Centre on Atypical Communication, Nijmegen, The Netherlands
- Stegenwallner-Schütz, M., & Adani, F. (2017). Numerusinformation vereinfacht das Satzverständnis: Querschnittsuntersuchungen zum Verständniserwerb von transitiven Sätzen mit Wortstellungsvariation. [Number dissimilarity facilitates sentence comprehension: Cross-sectional acquisition studies on the comprehension of transitive sentences with word order variation]. *LOGOS*, 25, 96–105.
- Stegenwallner-Schütz, M., & Adani, F. (2021). Number dissimilarity effects in object-initial sentence comprehension by German-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 64(3), 870–888. https://doi.org/10.1044/2020_JSLHR-19-00305
- Steinberg, L., Bornstein, M. H., Vandell, D. L., & Rook, K. (2011). *Life-span development: Infancy through adulthood*. Boston, MA: Wadsworth Cengage Learning.
- Stöhr, A., Akpinar, D., Bianchi, G., & Kupisch, T. (2012). Gender marking in Italian–German heritage speakers and L2-learners of German. In K. Braunmüller & C. Gabriel (Eds.), *Multilingual Individuals Multilingual Societies (MIMS)* (pp. 153–170). Amsterdam: John Benjamins.
- Stokes, S. F., & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry*, 50(4), 498–505. <https://doi.org/10.1111/j.1469-7610.2008.01991.x>
- Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech, and Hearing Services in Schools*, 48(1), 16–30. https://doi.org/10.1044/2016_LSHSS-16-0014
- Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: a follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, 41(2), 407–418. <https://doi.org/10.1044/jslhr.4102.407>
- Stow, C., & Dodd, B. (2005). A survey of bilingual children referred for investigation of communication disorders: a comparison with monolingual children referred in one area in England. *Journal of Multilingual Communication Disorders*, 3(1), 1–23. <https://doi.org/10.1080/14769670400009959>
- Strik, N. (2009). Derivational contrasts in Dutch and French: Evidence from the acquisition of Long Distance Wh-questions. *Linguistics in the Netherlands*, 26(1), 91–102. <https://doi.org/10.1075/avt.26.09str>
- Strik, N., & Pérez-Leroux, A. T. (2011). Jij doe wat girafe?: Wh-movement and inversion in Dutch-French bilingual children. *Linguistic Approaches to Bilingualism*, 1(2), 175–205. <https://doi.org/10.1075/lab.1.2.03str>
- Stromswold, K. (1995). The acquisition of subject and object Wh-questions. *Language Acquisition: A Journal of Developmental Linguistics*, 4(1-2), 5–48. https://doi.org/10.1207/s15327817la0401&2_1
- Stromswold, K. (2001). The Heritability of Language: A Review and Metaanalysis of Twin, Adoption, and Linkage Studies. *Language*, 77(4), 647–723. <https://doi.org/10.1353/lan.2001.0247>
- Sua-Kay, E., & Santos, M. E. (2014). *Grelha de Avaliacao da Linguagem – nível escolar (GOL-E)*, 2nd ed. Alcoitao: Escola Superior de Saúde do Alcoitao.
- Sutherland, D., & Gillon, G. T. (2005). Assessment of phonological representations in children with speech impairment. *Language, Speech, and Hearing Services in Schools*, 36(4), 294–307. [https://doi.org/10.1044/0161-1461\(2005/030\)](https://doi.org/10.1044/0161-1461(2005/030))
- Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical

- problem solving. *Journal of Educational Psychology*, 98(2), 265–281. <https://doi.org/10.1037/0022-0663.98.2.265>
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1(1), 1–26. <https://doi.org/10.1111/1529-1006.001>
- Szterman, R., & Friedmann, N. (2015). Insights into the syntactic deficit of children with hearing impairment from a sentence repetition task. In C. Hamann & E. Ruigendijk (Eds.), *Language acquisition and development: Generative approaches to language acquisition* (pp. 492–505). Newcastle: Cambridge Scholars Publishing.
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. <https://doi.org/10.1017/S0954394512000129>
- Taha, J., Stojanovik, V., & Pagnamenta, E. (2021a). Expressive Verb Morphology Deficits in Arabic-Speaking Children With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, 64(2), 561–578. https://doi.org/10.1044/2020_JSLHR-19-00292
- Taha, J., Stojanovik, V., & Pagnamenta, E. (2021b). Sentence Repetition as a Clinical Marker of Developmental Language Disorder: Evidence From Arabic. *Journal of Speech, Language, and Hearing Research*, 64(12), 4876–4899. https://doi.org/10.1044/2021_JSLHR-21-00244
- Talli, I., & Stavrakaki, S. (2020). Short-term memory, working memory and linguistic abilities in bilingual children with Developmental Language Disorder. *First Language*, 40(4), 437–460. <https://doi.org/10.1177/0142723719886954>
- Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology*, 107(3), 306–324. <https://doi.org/10.1016/j.jecp.2010.05.006>
- Tamburelli, M., & Jones, G. (2013). Investigating the relationship between nonword repetition performance and syllabic structure in typical and atypical language development. *Journal of Speech, Language, and Hearing Research*, 56(2), 708–720. [https://doi.org/10.1044/1092-4388\(2012/11-0171\)](https://doi.org/10.1044/1092-4388(2012/11-0171))
- Tamburelli, M., Sanoudaki, E., Jones, G., & Sowinska, M. (2015). Acceleration in the bilingual acquisition of phonological structure: Evidence from Polish–English bilingual children. *Bilingualism: Language and Cognition*, 18(4), 713–725. <https://doi.org/10.1017/S1366728914000716>
- Tempel, G., Gerken, C., & Sadowski, G. (2018). *Gesund in der Schule: Ergebnisse der Schuleingangsuntersuchung 2016/2017 in Bremen*. https://www.gesundheitsamt.bremen.de/detail.php?template=20_search_d&search%5Bsend%5D=true&lang=de&search%5Bvt%5D=schuleingangsuntersuchung
- Thal, D., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research*, 42(2), 482–496. <https://doi.org/10.1044/jslhr.4202.482>
- Thatcher, K. L. (2010). The development of phonological awareness with specific language-impaired and typical children. *Psychology in the Schools*, 47(5), 467–480. <https://doi.org/10.1002/pits.20483>
- Thelen, K. (2014). Störungen der Grammatik zwischen 3;0 und 5;0 Jahren. In A. Fox-Boyer (Ed.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen- Kindergartenphase* (pp. 55–72). München: Elsevier/Urban & Fischer.
- Thomas, E. M., Williams, N., Jones, L. A., Davies, S., & Binks, H. (2014). Acquiring complex structures under minority language conditions: Bilingual acquisition of plural morphology in Welsh. *Bilingualism: Language and Cognition*, 17(3), 478–494. <https://doi.org/10.1017/S1366728913000497>
- Thompson, C. K., Tait, M. E., Ballard, K. J., & Fix, S. C. (1999). Agrammatic aphasic subjects' comprehension of subject and object extracted Wh questions. *Brain and Language*, 67(3), 169–187. <https://doi.org/10.1006/brln.1999.2052>
- Theodorou, E., Kambanaros, M., & Grohmann, K. K. (2017). Sentence repetition as a tool for screening morphosyntactic abilities of bilingual children with SLI. *Frontiers in Psychology*, 8,

2104. <https://doi.org/10.3389/fpsyg.2017.02104>
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426-445. <http://dx.doi.org/10.1177/1367006911403202>
- Thordardottir, E. (2012). *Proposed Diagnostic Procedures and Criteria for Bi-SLI COST Action Studies*. Available at: <http://www.bi-sli.org>
- Thordardottir, E. (2014). The typical development of simultaneous bilinguals: Vocabulary, morphosyntax and language processing in two age groups of Montreal preschoolers. In T. Grütter & J. Paradis (Eds.), *Input and Experience in Bilingual Development* (pp. 141-160). Amsterdam: John Benjamins.
- Thordardottir, E. (2015a). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 331-358). Bristol: Multilingual Matters.
- Thordardottir, E. (2015b). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 17, 97-114. <http://dx.doi.org/10.3109/17549507.2014.923509>
- Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, 23(1), 236-255. <https://doi.org/10.1177/13670069177224>
- Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46(1), 1-16. <https://doi.org/10.1016/j.jcomdis.2012.08.002>
- Thordardottir, E. T., & Juliusdottir, A. G. (2013). Icelandic as a second language: a longitudinal study of language knowledge and processing by school-age children. *International Journal of Bilingual Education and Bilingualism*, 16(4), 411-435. <https://doi.org/10.1080/13670050.2012.693062>
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, 54(2), 580-597. [https://doi.org/10.1044/1092-4388\(2010/09-0196\)](https://doi.org/10.1044/1092-4388(2010/09-0196))
- Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4(1), 1-21. <https://doi.org/10.1080/14769670500215647>
- Tibshirani, R., Guenther, W., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423. <https://doi.org/10.1111/1467-9868.00293>
- Tomas, E., Demuth, K., Smith-Lock, K. M., & Petocz, P. (2015). Phonological and morphophonological effects on grammatical development in children with specific language impairment. *International Journal of Language & Communication Disorders*, 50(4), 516-528. <https://doi.org/10.1111/1460-6984.12152>
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tomblin, J. B. (1989). Familial concentration of developmental language impairment. *Journal of Speech & Hearing Disorders*, 54(2), 287-295. <https://doi.org/10.1044/jshd.5402.287>
- Tomblin, J. B. (2008). Validating diagnostic standards for specific language impairment using adolescent outcomes. In C. F. Norbury, J. B. Tomblin & D. V. M. Bishop (Eds.), *Understanding developmental language disorders: From theory to practice* (pp. 93-117). Hove, UK: Psychology Press.
- Tomblin, J. B. (2015). Children with Specific Language Impairment (SLI). In E. Bavin & L. Naigles (Eds.), *The Cambridge Handbook of Child Language* (pp. 527-544). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781316095829.024>
- Tomblin, J. B., Freese, P. R., & Records, N. L. (1992). Diagnosing specific language impairment in adults for the purpose of pedigree analysis. *Journal of Speech and Hearing Research*, 35(4), 832-843. <https://doi.org/10.1044/jshr.3504.832>

- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245–1260. <https://doi.org/10.1044/jslhr.4006.1245>
- Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39*(6), 1284–1294. <https://doi.org/10.1044/jslhr.3906.1284>
- Tomblin, J. B., & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders* (pp. 361–382). The MIT Press.
- Topbaş, S., & Güven, S. (2011). *Türkçe erken dil gelişim testi-TEDİL-3 [Test of Early Language Development: TELD-3, Turkish Version]*. Ankara: Detay Yayıncılık.
- Torrens, V. (1995). The acquisition of inflection in Spanish and Catalan. In C. T. Schuetze, J. B. Ganger & K. Broihier (Eds.), *Papers on language processing and acquisition (MITWPL 26)* (pp. 451–472). Cambridge, MA: MIT Working Papers in Linguistics.
- Tracy, R. (1986). The acquisition of case morphology in German. *Linguistics, 24*(1), 47–78. <https://doi.org/10.1515/ling.1986.24.1.47>
- Tracy, R. (1991). *Sprachliche Strukturentwicklung. Linguistische und kognitionspsychologische Aspekte des Erstspracherwerbs*. Tübingen: Gunter Narr.
- Tracy, R. (1994). Raising Questions: Formal and Functional Aspects of the Acquisition of WÄ-Questions in German. In R. Tracy & R. Lattey (Eds.) *How Tolerant Is Universal Grammar: Essays on Language Learnability and Language Variation* (pp. 1-34). Tübingen: Niemeyer.
- Tracy, R., & Thoma, D. (2009). Convergence on finite V2 clauses in L1, bilingual L1 and early L2 acquisition. In C. Dimroth & P. Jordens (Eds.), *Functional Categories in Learner Language* (pp. 1-44). Berlin: De Gruyter.
- Trauner, D., Wulfeck, B., Tallal, P., & Hesselink, J. (1995). *Neurologic and MRI profiles of language impaired children* (Technical Report CND-9513). Center for Research in Language, University of California at San Diego
- Trauner, D., Wulfeck, B., Tallal, P., & Hesselink, J. (2000). Neurological and MRI profiles of children with developmental language impairment. *Developmental medicine and child neurology, 42*(7), 470–475. <https://doi.org/10.1017/s0012162200000876>
- Travis, L. (1991). Parameters of phrase structure and verb second phenomena. In R. Freidin (Ed.), *Principles and Parameters in Comparative Grammar* (pp. 339–364). Cambridge, MA: MIT Press.
- Tsimpli, I. (2014). Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism, 4*(3), 283–313. <https://doi.org/10.1075/lab.4.3.01tsi>
- Tsimpli, I., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research, 23*, 215-242.
- Tsimpli, I. M., & Mastropavlou, M. (2008). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 143–183). Mahwah, NJ: Lawrence Erlbaum.
- Tucci, I., & Groh-Samberg, O. (2008). Das enttäuschte Versprechen der Migration: Migrantennachkommen in Frankreich und Deutschland. *Schweizerische Zeitschrift für Soziologie, 34*(2), 307–334.
- Tuller, L. (2015). Clinical use of parental questionnaires in multi-lingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). Multilingual Matters. <https://doi.org/10.21832/9781783093137-013>
- Tuller, L., Abboud, L., Ferré, S., Fleckstein, A., Prévost, P., dos Santos, C., et al. (2015). Specific language impairment and bilingualism: assembling the pieces. In C. Hamann & E. Ruigendijk (Eds.), *Language Acquisition and Development. Proceedings of GALA 2013* (pp. 533-567). Newcastle: Cambridge Scholars Press.
- Tuller, L., Delage, H., Monjauze, C., Piller, A. G., & Barthez, M. A. (2011). Clitic pronoun production as a measure of atypical language development in French. *Lingua, 121*(3), 423–41. <https://doi.org/10.1016/j.lingua.2010.10.008>

- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Abed Ibrahim, L., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, 53(4), 888–904. <https://doi.org/10.1111/1460-6984.12397>
- Tuller, L., Henry, C., Sizaret, E., & Barthez, M.-A. (2012). Specific language impairment at adolescence: Avoiding complexity. *Applied Psycholinguistics*, 33(1), 161–184. <https://doi.org/10.1017/S0142716411000312>
- Tuppat, J., Becker, B., & Lange, M. (2016). The risk of delayed school entry in relation to migration background in different institutional settings. *Zeitschrift für Erziehungswissenschaft*, 19(4), 745–764.
- Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language*, 4(2), 211–224. <https://doi.org/10.1017/S0305000900001616>
- Ullman M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews. Neuroscience*, 2(10), 717–726. <https://doi.org/10.1038/35094573>
- Ullman M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231–270. <https://doi.org/10.1016/j.cognition.2003.10.008>
- Ullman, M. T. (2013). The role of declarative and procedural memory in disorders of language. *Linguistic Variation*, 13(2), 133–154. <https://doi.org/10.1075/lv.13.2.01ull>
- Ullman, M., & Gopnik, M. (1994). Past tense production: Regular, irregular and nonsense verbs. *McGill Working Papers in Linguistics*, 10, 81–118.
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 41(3), 399–433. [https://doi.org/10.1016/S0010-9452\(08\)70276-4](https://doi.org/10.1016/S0010-9452(08)70276-4)
- Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience and Biobehavioral Reviews*, 51, 205–222. <https://doi.org/10.1016/j.neubiorev.2015.01.008>
- Umemeke Chikere, C. M., Wilson, K. J., Allen, A. J., & Vale, L. (2021). Comparative diagnostic accuracy studies with an imperfect reference standard - a comparison of correction methods. *BMC Medical Research Methodology*, 21(1), 67. <https://doi.org/10.1186/s12874-021-01255-4>
- Ünsal, F., & Fox, A. V. (2002). Lautspracherwerb bei zweisprachigen Migrantenkindern (Türkisch–Deutsch). *Forum Logopädie*, 16(3), 10–15.
- Unsworth, S. (2005). Child L2, adult L2, child L1 differences and similarities. A study on the acquisition of object scrambling in Dutch. *Doctoral dissertation*. Utrecht Institute of Linguistics OTS, LOT, Netherlands Graduate School of Linguistics.
- Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 365–396. <https://doi.org/10.1177/0267658308090185>
- Unsworth, S. (2013a). Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, 33(3), 21–50.
- Unsworth, S. (2013b). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language & Cognition*, 16, 86–110. <http://dx.doi.org/10.1017/S1366728912000284>
- Unsworth, S. (2013c). Assessing age of onset effects in (early) child L2 acquisition. *Language Acquisition*, 20(2), 74–92. <https://doi.org/10.1080/10489223.2013.766739>
- Unsworth, S. (2014). Comparing the role of input in bilingual acquisition across domains. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 181–201). Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/tilar.13.10uns>
- Unsworth, S. (2015). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalán & J. Teffers-Daller (Eds.), *Language Dominance in Bilinguals* (pp.156–173). Cambridge: Cambridge University Press, <https://doi.org/10.1017/CBO9781107375345.008>
- Unsworth, S. (2016a). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of Child*

- Language*, 43(3), 603-634. <https://doi.org/10.1017/S030500091500080X>
- Unsworth, S. (2016b). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103–122). Mouton de Gruyter. <https://doi.org/10.1515/9783110341249-008>
- Unsworth, S. (2019). Quantifying experience in heritage language development. In M. Schmid & B. Köpcke (Eds.), *The Oxford Handbook of First Language Attrition* (pp. 434-445). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198793595.013.34>
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014a). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35(4), 765-805. <https://doi.org/10.1017/S0142716412000574>
- Unsworth, S., Brouwer, S., de Bree, E., & Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development: Degree of non-native input matters. *Applied Psycholinguistics*, 40(5), 1189-1219. <https://doi.org/10.1017/S0142716419000225>
- Unsworth, S., Chondrogianni, V., & Skarabela, B. (2018). Experiential measures can be used as a proxy for language dominance in bilingual language acquisition research. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2018.01809>
- Unsworth, S., Persson, L., Prins, T., & De Bot, K. (2014b). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 36(5), 527–548. <https://doi.org/10.1093/applin/amt052>
- Vaknin-Nusbaum, V., & Saiegh-Haddad, E. (2020). The contribution of morphological awareness to reading comprehension in Arabic-speaking second graders. *Reading and Writing*, 33(10), 2413-2436
- Valian, V. (2015). Bilingualism and cognition. *Bilingualism*, 18(1), 3–24. <https://doi.org/10.1017/S1366728914000522>
- van der Hulst, H., & van der Wijer, J. (1991). Topics in Turkish phonology. In H. E. Boeschoten & L. Verhoeven (Eds.), *Turkish linguistics today* (pp. 11-59). Leiden: E. J. Brill.
- van der Lely H. K. (1994). Canonical linking rules: forward versus reverse linking in normally developing and specifically language-impaired children. *Cognition*, 51(1), 29–72. [https://doi.org/10.1016/0010-0277\(94\)90008-6](https://doi.org/10.1016/0010-0277(94)90008-6)
- van der Lely, H.K.J. (1996). Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation. *Lingua*, 98(4), 243-272. [https://doi.org/10.1016/0024-3841\(95\)00044-5](https://doi.org/10.1016/0024-3841(95)00044-5)
- van der Lely, H.K.J. (1997a). Narrative discourse in Grammatical specific language impaired children: a modular language deficit?. *Journal of Child Language*, 24(1), 221–256. <https://doi.org/10.1017/s0305000996002966>
- van der Lely, H.K.J. (1997b). Language and cognitive development in a grammatical SLI boy: Modularity and innateness. *Journal of Neurolinguistics*, 10(2-3), 75–107. [https://doi.org/10.1016/S0911-6044\(97\)00011-0](https://doi.org/10.1016/S0911-6044(97)00011-0)
- van der Lely, H. K. J. (1998). SLI in Children: Movement, Economy, and Deficits in the Computational-Syntactic System. *Language Acquisition*, 7(2–4), 161–192. https://doi.org/10.1207/s15327817la0702-4_4
- van der Lely, H.K.J (2004). Evidence for and implications of a domain-specific grammatical deficit. In L. Jenkins (Ed.), *The Genetics of Language* (pp. 117–144). Oxford: Elsevier.
- van der Lely, H.K.J. (2005a). Grammatical-SLI and the computational grammatical complexity hypothesis. *Revue Frequences*, 17(3), 13-20.
- van der Lely, H.K.J. (2005b). Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends in cognitive sciences*, 9(2), 53–59. <https://doi.org/10.1016/j.tics.2004.12.002>
- van der Lely, H.K.J., & Battell, J. (2003). Wh-Movement in Children with Grammatical SLI: A Test of the RDDR Hypothesis. *Language*, 79(1), 153-181. <https://dx.doi.org/10.1353/lan.2003.0089>.
- van der Lely, H.K.J, Jones, M., & Marshall, C. R. (2011). Who did Buzz see someone? Grammaticality

- judgement of wh-questions in typically developing children and children with Grammatical-SLI. *Lingua*, 121(3), 408–422. <https://doi.org/10.1016/j.lingua.2010.10.007>
- van der Lely, H.K.J., & Stollwerck, L. (1997). Binding Theory and Specifically Language Impaired Children. *Cognition*, 62, 245–290. [https://doi.org/10.1016/S0010-0277\(96\)00783-4](https://doi.org/10.1016/S0010-0277(96)00783-4)
- van der Velde, M. (2003). *Les déterminants et les pronoms en néerlandais et en français: Analyse syntaxique et acquisition* [Doctoral Dissertation, Universities of Paris V & VIII].
- van der Velde, M., Jakubowicz, C., & Rigaut, C. (2002). The acquisition of determiners and pronominal clitics by three French-speaking children. In I. Lasser (Ed.), *The process of language acquisition* (pp. 115–132). Frankfurt: Peter Lang.
- van Dijk, C., van Wonderen, E., Koutamanis, E., Kootstra, G. J., Dijkstra, T., & Unsworth, S. (2022). Cross-linguistic influence in simultaneous and early sequential bilingual children: a meta-analysis. *Journal of Child Language*, 49(5), 897–929. <https://doi.org/10.1017/S0305000921000337>
- Vang Christensen, R., & Hansson, K. (2012). The use and productivity of past tense morphology in specific language impairment: An examination of Danish. *Journal of Speech, Language, and Hearing Research*, 55, 1671–1689.
- van Hout, A. (2000). Event semantics in the lexicon-syntax interface: Verb frame alternations in Dutch and their acquisition. In C. L. Tenny & J. Pustejovsky (Eds.), *Events as grammatical objects. The converging perspectives of lexical semantics and syntax* (pp. 239–282). Stanford, CA: CSLI.
- van Hout, A. (2008). Acquiring telicity cross-linguistically: On the acquisition of telicity entailments associated with transitivity. In M. Bowerman & P. Brown (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 255–278). New York, NY: Routledge.
- van Hout, A. (2018). On the acquisition of event culmination. In K. Syrett & S. Arunachalam (Eds.), *Semantics in Language Acquisition* (pp. 95–121). Amsterdam: John Benjamins. <https://doi.org/10.1075/tilar.24.05hou>
- Velnić, M. (2020). Acquisition of a transparent gender system: a comparison of Italian and Croatian. *Frontiers in Psychology*, 11, 571674. <https://doi.org/10.3389/fpsyg.2020.571674>
- Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, 141, 65–82. <https://doi.org/10.1016/j.jecp.2015.06.015>
- Verhoeven, L., Steenge, J., & van Balkom, H. (2011a). Verb morphology as clinical marker of specific language impairment: evidence from first and second language learners. *Research in developmental disabilities*, 32(3), 1186–1193. <https://doi.org/10.1016/j.ridd.2011.01.001>
- Verhoeven, L., Steenge, J., & van Balkom, H. (2012). Linguistic transfer in bilingual children with specific language impairment. *International Journal of Language & Communication Disorders*, 47(2), 176–183. <https://doi.org/10.1111/j.1460-6984.2011.00092.x>
- Verhoeven, L., Steenge, J., van Weerdenburg, M., & van Balkom, H. (2011b). Assessment of second language proficiency in bilingual children with specific language impairment: a clinical perspective. *Research in developmental disabilities*, 32(5), 1798–1807. <https://doi.org/10.1016/j.ridd.2011.03.010>
- Vernon-Feagans, L., Bratsch-Hines, M., Reynolds, E., & Willoughby, M. (2020). How Early Maternal Language Input Varies by Race and Education and Predicts Later Child Language. *Child development*, 91(4), 1098–1115. <https://doi.org/10.1111/cdev.13281>
- Vernon-Feagans, L., Mokrova, I. L., Carr, R. C., Garrett-Peters, P. T., Burchinal, M. R., & Family Life Project Key Investigators (2019). Cumulative years of classroom quality from kindergarten to third grade: Prediction to children's third grade literacy skills. *Early Childhood Research Quarterly*, 47, 531–540. <https://doi.org/10.1016/j.ecresq.2018.06.005>
- Verrips, M., & Weissenborn, J. (1992). Routes to verb placement in early German and French: The independence of finiteness and agreement. In J. Meisel. (Ed.), *The Acquisition of Verb Placement: Functional Categories and V2 Phenomena in Language Development. (Studies in Theoretical Psycholinguistics)* (pp. 283–331). Dordrecht: Kluwer Academic Publishers.

- Victorino, K. R., & Schwartz, R. G. (2015). Control of Auditory Attention in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 58(4), 1245–1257. https://doi.org/10.1044/2015_JSLHR-L-14-0181
- Vikner, S. (1995). *Verb movement and expletive subjects in the Germanic languages*. Oxford: Oxford University Press.
- Vinther, T. (2002). Elicited imitation: a brief overview. *International Journal of Applied Psycholinguistics*, 12(1), 54–73. <https://doi.org/10.1111/1473-4192.00024>
- Voet Cornelli, B. (2020). *Die Qualität der Sprachentwicklungsdiagnostik bei mehrsprachigen Kindern—Pädiatrische Früherkennungsuntersuchungen auf dem Prüfstand*. Ph.D. Dissertation, University of Frankfurt, Frankfurt am Main, Germany.
- Voet Cornelli, B., Geist, B., Grimm, A., & Schulz, P. (2012). Praktiken der Sprachstandserhebung bei mehrsprachigen Kindern – erste Erkenntnisse aus dem Projekt cammino. In S. Jeuk, & J. Schäfer (Eds.), *Deutsch als Zweitsprache in Kindertageseinrichtungen und Schulen: Aneignung, Förderung, Unterricht. Beiträge aus dem 7. Workshop Kinder mit Migrationshintergrund* (pp 43–74). Freiburg im Breisgau: Fillibach.
- Voet Cornelli, B., Schulz, P., & R. Tracy (2013). Sprachentwicklungsdiagnostik bei Mehrsprachigkeit: Eine Herausforderung für die pädiatrische Praxis. *Monatsschrift Kinderheilkunde*, 161, 911–917. <https://doi.org/10.1007/s00112-012-2752-z>
- Volkers, N. (2018). Diverging views of language disorders: Researchers debate whether the label “developmental language disorder” should replace “specific language impairment.” *The ASHA Leader*, 23(12), 44–53. <https://doi.org/10.1044/leader.FTR1.23122018.44>
- von Stechow, A. (1995). On the proper treatment of tense. *Semantics and Linguistic Theory*, 5, 362–386. <https://doi.org/10.3765/salt.v5i0.2700>
- von Stechow, A., & Sternefeld, W. (1988). *Bausteine syntaktischen Wissens. Ein Lehrbuch der generativen Grammatik*. Opladen: Westdeutscher Verlag.
- Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: a meta-analysis. *Research in developmental disabilities*, 34(9), 2586–2597. <https://doi.org/10.1016/j.ridd.2013.05.014>
- Vugs, B., Hendriks, M., Cuperus, J., Knoors, H., & Verhoeven, L. (2017). Developmental Associations Between Working Memory and Language in Children With Specific Language Impairment: A Longitudinal Study. *Journal of Speech, Language, and Hearing Research*, 60(11), 3284–3294. https://doi.org/10.1044/2017_JSLHR-L-17-0042
- Vugs, B., Hendriks, M., Cuperus, J., & Verhoeven, L. (2014). Working memory performance and executive function behaviors in young children with SLI. *Research in developmental disabilities*, 35(1), 62–74. <https://doi.org/10.1016/j.ridd.2013.10.022>
- Vugs, B., Knoors, H., Cuperus, J., Hendriks, M., & Verhoeven, L. (2016). Interactions between working memory and language in young children with specific language impairment (SLI). *Child Neuropsychology: a journal on normal and abnormal development in childhood and adolescence*, 22(8), 955–978. <https://doi.org/10.1080/09297049.2015.1058348>
- Vuksanović, J., Avramović-Ilić, I., Abramović, I., & Bjekic, J. (2012). Razumevanje svojstva iscrpnosti u višestrukim pitanjima kod predškolske dece tipičnog razvoja [Comprehension of exhaustive wh-questions of typically developing preschool children]. *Zbornik Instituta za pedagoška istraživanja*, 44(2), 368–384. <https://doi.org/10.2298/ZIPII202368V>
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192–212. <https://doi.org/10.1037/0033-2909.101.2.192>
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: Pro-Ed.
- Waltersbacher, A. (2016). *Heilbericht 2016 des Wissenschaftlichen Instituts der AOK (WIdO)*. https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/Buchreihen/Heilmittelbericht/wido_hei_hmb_2016.pdf

- Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: differentiating typical and impaired language learners. *Journal of Speech and Hearing Research*, 38(6), 1349–1355. <https://doi.org/10.1044/jshr.3806.1349>
- Watson, J. C. E. (2007). *The phonology and morphology of Arabic. The phonology of the world's languages*. Oxford University Press.
- Watson, J. C. E. (2011). Arabic dialects (general article). In S. Weninger (Ed.), *The Semitic languages* (Vol. 36, pp. 851–896). De Gruyter. <https://doi.org/10.1515/9783110251586.851>
- Weber, R. C., Johnson, A., Riccio, C. A., & Liew, J. (2016). Balanced bilingualism and executive functioning in children. *Bilingualism: Language and Cognition*, 19(2), 425–431. <https://doi.org/10.1017/S1366728915000553>
- Weber-Fox, C., & Neville, H. (1999). Functional neural subsystems are differentially affected by delays in second language immersion: ERP and behavioral evidence in bilinguals. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 23–38). Lawrence Erlbaum.
- Weber-Fox, C., & Neville, H. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of bilinguals. *Journal of Speech Language and Hearing Research*, 44(6), 1338–1353. [https://doi.org/10.1044/1092-4388\(2001/104\)](https://doi.org/10.1044/1092-4388(2001/104))
- Weckerly, J., Wulfeck, B., & Reilly, J. (2001). Verbal fluency deficits in children with specific language impairment: slow rapid naming or slow to name? *Child Neuropsychology*, 7, 142–152.
- Wegener, H. (1992). *Kindlicher Zweitspracherwerb. Untersuchungen zur Morphologie des Deutschen und ihrem Erwerb durch Kinder mit polnischer, russischer und türkischer Erstsprache. Eine Längsschnittuntersuchung*. [Unpublished habilitation thesis, University of Augsburg], Germany.
- Wegener, H. (1994). Variation in the acquisition of German plural morphology by second language learners. In R. Tracy & E. Lattey (Eds.), *How tolerant is universal grammar? Essays on language learnability and language variation* (pp. 267–294). Max Niemeyer Verlag. <https://doi.org/10.1515/9783111634777.267>
- Weiss, A., Leonard, L., Rowan, L., & Chapman, K. (1983). Linguistic and non-linguistic features of style in normal and language-impaired children. *Journal of Speech and Hearing Disorders*, 48, 154–164.
- Weissenborn, J. (1990). Functional categories and verb movement in early German: The acquisition of German syntax reconsidered. In M. Rothweiler (Ed.), *Language acquisition and grammar: Linguistic studies on the acquisition of syntax and morphology (Linguistische Berichte Sonderheft 3)* (pp. 190–224). Opladen: Westdeutscher Verlag.
- Weissenborn, J. (1991). Null Subjects in Early Grammars. In J. Weissenborn, H. Goodluck, & T. Roeper (Eds.), *Theoretical Issues in Language Acquisition* (pp. 269–299). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Weissenborn, J. (2000). Der Erwerb von Morphologie und Syntax. In H. Grimm (Ed.), *Sprachentwicklung* (pp. 141–169). Göttingen: Hogrefe.
- Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, 37(2), 278–286. <https://doi.org/10.1037/h0080725>
- Wexler, K. (1994). Optional Infinitives, head movement and the economy of derivations. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 305–350). Cambridge: Cambridge University Press.
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua*, 106(1-4), 23–79. [https://doi.org/10.1016/S0024-3841\(98\)00029-1](https://doi.org/10.1016/S0024-3841(98)00029-1)
- Wexler, K., Schaeffer, J., & Bol, G. (2004). Verbal syntax and morphology in typically developing Dutch children and children with SLI: How developmental data can play an important role in morphological theory. *Syntax*, 7(2), 148–198. <https://doi.org/10.1111/j.1467-9612.2004.00006.x>
- Wexler, K., Schütze, C. T., & Rice, M. (1998). Subject Case in Children With SLI and Unaffected Controls: Evidence for the Agr/Tns Omission Model. *Language Acquisition*, 7(2–4), 317–344. https://doi.org/10.1207/s15327817la0702-4_8
- White, L. (1991). Adverb placement in second language acquisition: some effects of positive and negative

- evidence in the classroom. *Interlanguage Studies Bulletin (Utrecht)*, 7(2), 133-161. <https://doi.org/10.1177/02676583910070020>
- White, L. (2003). *Second language acquisition and Universal Grammar*. New York: Cambridge University Press.
- White, L. (2018). Formal linguistics and second language acquisition. In D. Miller, F. Bayram, J. Rothman & L. Serratrice, *Bilingual Cognition and Language: The state of science across its subfields* (pp. 57-80). Amsterdam: John Benjamins.
- White, L. (2020). Linguistic Theory, Universal Grammar, and Second Language Acquisition. In B. VanPatten, G. D. Keating & S. Wulff (Eds.), *Theories in Second Language Acquisition (3rd ed.)* (pp. 34-53). New York: Routledge.
- White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, 12(3), 233-265. <https://doi.org/10.1177/026765839601200>
- Whitehouse, A., Line, E., Watt, H., & Bishop, D. V. M. (2009). Qualitative aspects of developmental language impairment relate to language and literacy outcome in adulthood. *International Journal of Language & Communication Disorders*, 44(4), 489-510. <https://doi.org/10.1080/13682820802708080>
- Whong-Barr, M., & Schwartz, B. D. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition*, 24, 579-616. <https://doi.org/10.1017/S0272263102004035>
- Widmer, M., Auderset, S., Nichols, J., Widmer, P., & Bickel, B. (2017). NP recursion over time: Evidence from Indo-European. *Language*, 93(4), 799-826. <https://doi.org/10.1353/lan.2017.0058>
- Wiefferink, K., van Beugen, C., Wegener Sleswijk, B., & Gerrits, E. (2020). Children with language delay referred to Dutch speech and hearing centres: caseload characteristics. *International Journal of Language & Communication Disorders*, 55(4), 573-582. <https://doi.org/10.1111/1460-6984.12540>
- Wiese, R. (1991). Was ist extrasilbisch im Deutschen und warum? *Zeitschrift für Sprachwissenschaft*, 10(1), 112-133. <https://doi.org/10.1515/zfs.1991.10.1.112>
- Wiese, R. (1996). *The phonology of German*. New York: Oxford University Press.
- Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.
- Wiig, E. H., Semel, E., & Secord, W. A. (2020). *Clinical Evaluation of Language Fundamentals—Fifth Edition. Deutsche Fassung*. Frankfurt: Pearson.
- Williams, K. T. (2007). *The Expressive Vocabulary Test* (2nd ed.). Circle Pines, MN: AGS Publishing.
- Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology*, 14(3), 292-305. <https://doi.org/10.3109/17549507.2011.636071>
- Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, 9(4-6), 349-363. <https://doi.org/10.1080/09658210143000155>
- Wimmer, E., & Scherger, A. L. (2022). Working Memory Skills in DLD: Does Bilingualism Make a Difference? *Languages*, 7(4), 287. <https://doi.org/10.3390/languages7040287>
- Windsor, J., Kohnert, K., Lobitz, K., & Pham, G. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, 19(4), 298-310. [https://doi.org/10.1044/1058-0360\(2010/09-0064\)](https://doi.org/10.1044/1058-0360(2010/09-0064))
- Winstanley, M., Durkin, K., Webb, R. T., & Conti-Ramsden, G. (2018a). Financial capability and functional financial literacy in young adults with developmental language disorder. *Autism & Developmental Language Impairments*, 3, 1-11. <https://doi.org/10.1177/2396941518794500>
- Winstanley, M., Webb, R. T., & Conti-Ramsden, G. (2018b). More or less likely to offend? Young adults with a history of identified developmental language disorders. *International Journal of Language & Communication Disorders*, 53(2), 256-270. <https://doi.org/10.1111/1460-6984.12339>
- Wode, H. (1983). On the systematicity of L1 transfer in L2 acquisition. In H. Wode (Ed.), *Papers on*

- language acquisition, language learning and language teaching* (pp. 144–149). Heidelberg: J. Groos
- Wojtecka, M. (2020). *Acquisition Pace and Developmental Path of Early Second Language Learners of German. A Longitudinal Study on Acquisition of Morphosyntax and Semantics* [Doctoral dissertation, Johann Wolfgang Goethe-Universität Frankfurt am Main].
- Wojtecka, M., Koch, C., Grimm, A., & Schulz, P. (2011). Production and comprehension of sentence negation in child German. In A. Grimm, A. Müller, C. Hamann & E. Ruigendijk (Eds.), *Production-Comprehension Asymmetries in Child Language* (pp. 217–246). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110259179.217>
- Wojtecka, M., & Schulz, P. (2011). Exhaustive wh-questions in eL2TD and SLI children. *Paper presented at the Eskisehir COST meeting, May 2011*
- Wojtecka, M., Schwarze, R., Grimm, A., & Schulz, P. (2013). Finiteness and verb placement in German: A challenge for early second language learners? In J. Cabrelli Amaro, T. Judy & D. Pascual y Cabo (Eds.), *Proceedings of the 12th Generative Approaches to Second Language Acquisition Conference (GASLA 2013)* (pp. 211–221). Somerville: Cascadia Press.
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91(3), 415–438. <https://doi.org/10.1037/0022-0663.91.3.415>
- Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, 14(3), 412–422. <https://doi.org/10.1017/S1366728910000611>
- Yavuz, H., Balci, A. (2011). Turkish Phonology and Morphology. In Z. Balpinar (Ed.), *Turkish Phonology and Morphology*. Eskisehir: Anadolu Universitesi
- Yew, S.G.K., & O’Kearney, R. (2013). Emotional and behavioral outcomes later in childhood and adolescence for children with specific language disorders: Meta-analyses of controlled prospective studies: SLI and emotional and behavioural disorders. *Journal of Child Psychology and Psychiatry*, 54(5), 516–524. <https://doi.org/10.1111/jcpp.12009>
- Yip, V., & Matthews, S. (2000). Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition*, 3(3), 193–208. <https://doi.org/10.1017/S136672890000033>
- Yip, V., & Matthews, S. (2007). Relative clauses in Cantonese-English bilingual children: Typological challenges and processing motivations. *Studies in Second Language Acquisition*, 29(2), 277–300. <https://doi.org/10.1017/S0272263107070143>
- Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology*, 2, 210. <https://doi.org/10.3389/fpsyg.2011.00210>
- Zaretsky, E., & Bar-Shalom, E. G. (2010). Does reading in shallow L1 orthography slow attrition of language-specific morphological structures? *Clinical Linguistics & Phonetics*, 24, 401–415. <http://dx.doi.org/10.3109/02699200903532532>
- Zebib, R., Henry, G., Khomsi, A., Messara, C., & Hreich, E. (2017). *Batterie d’Evaluation du Langage Oral chez l’enfant libanais (ELO-L)*. Kerserwan: LTE.
- Zebib, R., Henry, G., Messarra, C., Hreich, E. K., & Khomsi, A. (2019). ELO-L: A Norm-Referenced Language Screening Test for 3 to 8-Year-Old Lebanese Children. *Arab Journal of Applied Linguistics*, 4(2), 24–53. <https://arjals.com/ajal/article/view/146>
- Zebib, R., Prévost, P., Tuller, L., & Henry, G. (Eds.) (2020a). *Plurilinguisme et Troubles Spécifiques du Langage au Liban*. Beyrouth: Presses universitaires de l’Université Saint Joseph. <https://de.calameo.com/read/0061189916ca7c0fbbc03>
- Zebib, R., Tuller, L., Hamann, C., Abed Ibrahim, L., & Prévost, P. (2020b). Syntactic complexity and verbal working memory in bilingual children with and without Developmental Language Disorder. *First Language*, 40(4), 461–484. <https://doi.org/10.1177/0142723719888372>
- Zdorenko, T., & Paradis, J. (2008). The acquisition of articles in child L2 English: fluctuation, transfer, or both? *Second Language Research*, 24, 227–250. <https://doi.org/10.1177/0267658307086302>
- Zdorenko, T., & Paradis, J. (2012). Articles in child L2 English: When L1 and L2 acquisition meet at the

- interface. *First Language*, 32(1-2), 38-62. <https://doi.org/10.1177/01427237103967>
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., & Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68(3), vii-137. <https://doi.org/10.1111/j.0037-976x.2003.00260.x>
- Zesiger, P., Zesiger, L. C., Arabatzi, M., Baranzini, L., Cronel-Ohayon, S., Franck, J., ... Rizzi, L. (2010). The acquisition of pronouns by French children: A parallel study of production and comprehension. *Applied Psycholinguistics*, 31(4), 571-603. <https://doi.org/10.1017/S0142716410000147>
- Ziethe, A., Eysholdt, U., & Doellinger, M. (2013). Sentence repetition and digit span: Potential markers of bilingual children with suspected SLI? *Logopedics Phoniatrics Vocology*, 38(1), 1-10. <https://doi.org/10.3109/14015439.2012.664652>
- Zobl, H. (1980). The formal and developmental selectivity of L1 influence in L2 acquisition. *Language Learning*, 30(1), 43-57. <https://doi.org/10.1111/j.1467-1770.1980.tb00150.x>
- Zurer-Pearson, B. (2010). We can no longer afford a monolingual norm. *Applied Psycholinguistics*, 31(2), 339-343. <https://doi.org/10.1017/S014271640999052X>
- Zurer-Pearson, B., & Amaral, L. (2014). Interactions between input factors in bilingual language acquisition. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp.99-117). Amsterdam: John Benjamins.
- Zwart, J. W. (1993). *Dutch Syntax* [Doctoral dissertation, University of Groningen].
- Zwart, J. W. (1997). *Morphosyntax of Verb Movement. A Minimalist Approach to the Syntax of Dutch*. Dordrecht: Kluwer Academic Publishers.

Appendices

I Items of the German LITMUS-SRT

Level 1

S V agreement/ control for „werden“

Die Tomate wird morgen reif.	8 syll.	CW/nom-f
Der Himmel wird heute Nacht klar.	8 syll	CW/nom-m
Der böse Zauberer wird alt.	8 syll.	CW/nom-m

SVO agreement/case/ past

Die Cowboys fangen die Pferde	8 syll.	SVOpl/pres/acc-pl
Die Kinder suchen den Mann	7 syll.	SVOpl/pres/acc-m
Der kleine Hund bringt die Zeitung	8 syll.	SVOsg/pres/acc-f
Der Clown umarmte die Köchin	8 syll.	SVOsg/past/acc-f
Die Wikinger zeigten den Schatz	8 syll.	SVOpl/past/acc-m
Die Katze jagte die Maus	7 syll.	SVOsg/past/acc-f

Satzklammer (Sentence bracket – verb placement)

Die Clowns haben den Hund gesucht	9 syll.	SA/OV
Der Prinz hat die Prinzessin umarmt.	9 syll.	S A/OV
Die Köchin hat den Cowboy geweckt.	9 syll.	S A/O V
Der Bauer malt den Zauberer an.	9 syll.	S V O part
Der Roboter wirft die Tante um.	9 syll.	S V O part
Die Hexen machen das Paket auf	9 syll.	S V O part

Level 2**Object Wh questions with Wh-NP and with bare Wh**

Welchen Jungen mag der Zauberer?	9 syll.	O-Wh, Wh-NP
Welchen Bauern ärgert der Affe?	9 syll.	O-Wh, Wh-NP
Welchen Clown besucht der Wikinger?	9 syll.	O-Wh, Wh-NP
Wen umarmt der Pinguin heute?	9 syll.	O-Wh, bare Wh
Wen untersucht der Tierarzt gerade?	10 syll.	O-Wh, bare Wh
Wen beißt der große Löwe immer?	9 syll.	O-Wh, bare Wh

Biclausal sentences**Coordination**

Der Polizist jagt die Hexe und die Hexe lacht.	13 syll.	Coord Simp 3 DPs
Der König küsst die Königin und die Prinzessin weint.	14 syll.	Coord Simp 3 DPs
Der Prinz besucht den Ritter und die Königin malt.	13 syll.	Coord Simp 3 DPs

Non-finite complements

Die Königin will die starken Wikinger suchen	13 syll.	S /M O V
Die Prinzessin möchte den Zauberer aufwecken	13 syll.	S /M O V
Die Lehrerin möchte heute die Kinder malen	13 syll.	S/MOV

Complement-clauses/finite:

Der Clown möchte, dass die Frau den Mann besucht	11 syll.	Comp 2 DPS
Der Prinz will, dass der Ritter die Affen jagt.	11 syll.	Comp 3 DPs
Der Wikinger glaubt, dass die Hexe ihn mag.	11 syll.	Comp 2 DPs, pron

Level 3**Topicalization**

Den Clown umarmt der Wikinger am liebsten	11 syll.	Top (OVSA)
Den Arzt fotografiert der Bauer gerne.	11 syll.	Top (OVSA)
Den Koch besucht der Zauberer als ersten.	11 syll.	Top (OVSA)

Subject relatives

Ich sehe den Clown, der den Pinguin streichelt.	12 syll.	SR (2 DPs)
Ich sehe den Roboter, der den Cowboy weckt.	12 syll.	SR (2 DPs)
Ich sehe den Arzt, der den Affen untersucht.	12 syll.	SR (2 DPs)

Object relatives without an intervener

Der Roboter sieht den Mann, den er malen will.	12 syll.	OR, - int (2 DPs)
Die Frau sieht den Pinguin, den sie streicheln will.	12 syll.	OR, - int (2 DPs)
Die Hexe sieht den Raben, den sie gerne mag.	12 syll.	OR, - int (2 DPs)

Object relatives with an intervener

Ich sehe den Roboter, den der Bäcker malt.	12 syll.	OR, + int (2 DPs)
Ich sehe den Clown, den der Wikinger umarmt.	12 syll.	OR, + int (2 DPs)
Ich sehe den Vogel, den der Pinguin weckt.	12 syll.	OR, + int (2 DPs)

Long – reversible passives

Die Oma wird von dem großen Cowboy geärgert.	12 syll.	L-Pas
Der Clown wird von dem alten Roboter gefilmt.	12 syll.	L-Pas
Die nette Ärztin wird von der Mutter umarmt.	12 syll.	L-Pas

II Items of the German LITMUS-QU-NWR

Language-independent part		Language-dependent part	
Item	Item Type	Item	Item Type
pilu	LI control	sapi	LD control
kapi	LI control	Saku	LD control
lafi	LI control	kiS	LD control
faku	LI control	kas	LD control
pli	LI	Spaklu	LD
kla	LI	sfupli	LD
flu	LI	Spluki	LD
kip	LI	sklifu	LD
paf	LI	Splaklu	LD
fuk	LI	sklaplu	LD
paklu	LI	skifup	LD
fupli	LI	Spukif	LD
fluka	LI	skifapu	LD
plifu	LI	Spafika	LD
plaklu	LI	sfikupla	LD
flaplu	LI	skupifla	LD
kafip	LI	sklipafu	LD
pukif	LI	sflipuka	LD
kifapu	LI	Spiklafu	LD
pufaki	LI	skaflipu	LD
fikupla	LI	skapifuk	LD
kupafli	LI	Spifakup	LD
klipafu	LI	kafipS	LD
flipuka	LI	pukifs	LD
piklafu	LI	kapifaps	LD
kufłapi	LI	fikapuks	LD
kapufip	LI	pafluS	LD
pifakup	LI	kufłas	LD
flukif	LI	fikuspa	LD
klifak	LI	kufiski	LD
		pafiklas	LD
		fapuplaS	LD
		kluspi	LD
		plisfu	LD
		pusfa	LD
		faspi	LD

Notes: ‘S’ corresponds to the phoneme /ʃ/

III The LITMUS questionnaire for parents of bilingual children (German version)

Code des Kindes: _____ Datum des Elterngesprächs: _____

Name des Versuchsleiters: _____ Mutter ☐ Vater ☐

PaBiQ (BiLaD Elternfragebogen¹)

1. Allgemeine Informationen über das Kind:

1.1 Geburtsdatum: _____ Geburtsort/-land: _____

1.2 Wenn der Geburtsort nicht in Deutschland ist, Datum der Einreise:

1.3 Welche Sprachen spricht Ihr Kind zurzeit (mehr als ein paar isolierte Wörter)?

Arabisch / Portugiesisch / Türkisch /Englisch/ Andere	Deutsch	Andere	Andere

1.4 In welcher Sprache fühlt sich das Kind Ihrer Meinung nach am wohlsten ? _____

1.5 (Extrafrage) hat Ihr Kind Geschwister? Anzahl _____ Alter _____

2. Frühe Entwicklung des Kindes (Sprache und sonstige)

2.1 Wie alt war Ihr Kind, als es sein erstes Wort sprach (nicht Mama, Papa, bamababa
,(babbeln), sondern erkennbare Wörter etc.!)?

2.2 Wie alt war Ihr Kind, als es Wörter zusammenfügte, um erste kurze (Mini-) Sätze zu sagen?

Welche waren dies (möglichst lautgetreu aufschreiben!), zum Beispiel: *Bild malen; Nane essen;
mehr Milch; etc*

2.3 Haben Sie sich vor dem Alter von 3 – 4 Jahren einmal Sorgen über die Sprachentwicklung
Ihres Kindes gemacht? Hatten Sie damals Sorge, dass mit der Sprache Ihres Kindes etwas
nicht stimmt? JA NEIN

Kommentare (z.B. genauer fragen nach Diagnostik und Therapie):

2.4 Hatte Ihr Kind schon einmal Schwierigkeiten mit dem Hören oder häufige
Ohrentzündungen?

JA NEIN

Kommentare (hat Ihr Kind an den U-Untersuchungen teilgenommen? An welchen?):

2.5 In welchem Alter hatte Ihr Kind jeweils den ersten Kontakt zu seinen Sprachen ?

	Alter (Monate wenn möglich)		Alter (Monate wenn möglich)
Arabisch		Andere	
Deutsch		Andere	

2.6 Hatte Ihr Kind vor dem Alter von 4 Jahren generell Kontakt mit:

	0 Nie	1 Selten/ab und zu	2 Die Hälfte der Zeit	3 Oft	4 Sehr oft/ Immer
Arabisch					
Deutsch					
Andere					

2.7 In welchem Kontext und in welchem Alter fand dieser Kontakt (vor dem Alter von 4 Jahren) jeweils statt (Alter in jede zutreffende Zeile eintragen) ?

	Arabisch	Deutsch	Andere
a. Austausch mit der Mutter	Alter :	Alter :	Alter :
b. Austausch mit dem Vater	Alter :	Alter :	Alter :
c. Austausch mit den Großeltern	Alter :	Alter :	Alter :
d. Austausch mit der Babysitterin/ Haushaltshilfe	Alter :	Alter :	Alter :
e. Austausch mit anderen Erwachsenen (angeben)	Alter :	Alter :	Alter :
f. Austausch mit Geschwistern	Alter :	Alter :	Alter :
g. Kindergarten/Vorschule/Grundschule	Alter :	Alter :	Alter :
<i>Gesamt (1 Punkt pro Situation)</i>			
<i>Gesamt pro Sprache</i>	/7	/7	/7

3. Aktuelle Sprachfähigkeiten

	Arabisch	Deutsch	Andere/Dialekt
3.1 Wenn Sie Ihr Kind mit Kindern im gleichen Alter vergleichen, wie gut spricht es dann ...? <i>0 = nicht so gut/nicht so gut wie sie; 1 = ein bisschen weniger gut/fast so wie sie, aber ein kleines bisschen schlechter; 2 = eigentlich genau so gut; 3 = sehr gut/besser, viel besser</i>	0 1 2 3	0 1 2 3	0 1 2 3
3.2 Wenn Sie ihr Kind mit Kindern im gleichen Alter vergleichen, hat es Schwierigkeiten, korrekte Sätze zu bilden? <i>0 = nicht so gut/nicht so gut wie sie; bzw. große Schwierigkeiten 1 = ein bisschen weniger gut/ein kleines bisschen schlechter; 2 = eigentlich genau so gut; 3 = sehr gut/besser, keine Schwierigkeiten</i>	0 1 2 3	0 1 2 3	0 1 2 3
3.3 Sind Sie damit zufrieden, wie sich Ihr Kind im ausdrücken kann? Immer? <i>0 = überhaupt nicht zufrieden; 1 = nicht ganz zufrieden; 2 = ganz zufrieden; 3 = völlig zufrieden, sehr zufrieden</i>	0 1 2 3	0 1 2 3	0 1 2 3
3.4 Ist Ihr Kind unglücklich, wenn es auf nicht kommunizieren kann? <i>0 = sehr/fast immer unglücklich/frustriert; 1 = oft unglücklich /frustriert/ ja; 2 = manchmal, aber nicht oft; 3 = (fast)nie unglücklich/frustriert/nein</i>	0 1 2 3	0 1 2 3	0 1 2 3
3.5 Glauben Sie, dass Ihr Kind genauso gut spricht wie ein Kind im gleichen Alter, das nur...spricht? <i>0 = nicht so gut/nicht so gut wie sie; 1 = ein bisschen weniger gut/ein kleines bisschen schlechter/ fast wie sie mit ein paar Unterschieden ; 2 = eigentlich genau so gut; 3 = sehr gut/besser</i>	0 1 2 3	0 1 2 3	0 1 2 3
Gesamt nach Sprache	/15	/15	/15

4. Vergleich zwischen den Sprachen, die innerhalb der Familie verwendet werden¹⁷⁷

4.1 Mit den Eltern

Mutter ↔ Kind						Vater ↔ Kind				
	0 Nie	1 Selten/ ab und zu	2 Etwa die Hälfte der Zeit	3 Oft	4 Sehr oft/ Immer	0 Nie	1 Selten/ ab und zu	2 Etwa die Hälfte der Zeit	3 Oft	4 Sehr oft/ Immer
Arabisch										
Deutsch										
Andere Sprache :										

4.2 Mit einem anderen Erwachsenen, der regelmäßig auf Ihr Kind aufpasst (Opa, Oma, Babysitter etc.)

Anderer Erwachsener ↔ Kind					
	0 Nie	1 Selten/ ab und zu	2 Etwa die Hälfte der Zeit	3 Oft	4 Sehr oft/ Immer
Arabisch					
Deutsch					
Andere Sprache :					

Anmerkungen: _____

4.3 Sprachgebrauch mit den Geschwistern

Geschwister ↔ Kind					
	0 Nie	1 Selten/ ab und zu	2 Etwa die Hälfte der Zeit	3 Oft	4 Sehr oft/ Immer
Arabisch					
Deutsch					
Andere Sprache					

*In der Gegenwart der Eltern sprechen die Kinder Arabisch untereinander. Wenn die Eltern nicht dabei sind, neigen sie dazu, Deutsch zu benutzen.

¹⁷⁷ Wenn die Eltern fragen, wie sie die Wörter „selten“, etc., interpretieren sollen, kann man die folgende Erklärung geben: „Von der Gesamtzeit des sprachlichen Austausches zwischen Ihnen/seinem Vater/den Freunden Ihrer Familie und Ihrem Kind, wieviel Zeit davon ist auf Sprache X und wieviel Zeit auf Deutsch? Auf 4-Punkte Skala (wenn 0 für „nie“ steht und 4 für „oft/immer“), würden Sie sagen, dass Sprache X 4 ist und Deutsch 0, oder 3 Sprache X und 1 Deutsch, oder 2 Sprache X und 2 Deutsch, oder 1 Sprache X und 3 Deutsch, oder 0 Sprache X und 4 Deutsch?“

5. Sprachgebrauch in anderen Kontexten

5.1 Welche der folgenden Aktivitäten macht Ihr Kind jede Woche und in welcher (-n) Sprache(n)?

	Arabisch			Deutsch			Andere Sprache: Englisch		
Aktivitäten	0 Nie oder fast nie	1 Wenigstens einmal in der Woche	2 Jeden Tag	0 Nie oder fast nie	1 Wenigstens einmal in der Woche	2 Jeden Tag	0 Nie oder fast nie	1 Wenigstens einmal in der Woche	2 Jeden Tag
a. Lesen (Bücher, Magazine, Comics, Zeitungen, Zeitschriften)									
b. Fernsehen/Filme/Kino									
c. Geschichten erzählen/erzählt bekommen									
Gesamt									
Gesamt nach Sprache	/6			/6			/6		

5.2 Welche Sprache(n) spricht Ihr Kind mit den Freunden/Freundinnen, mit denen es regelmäßig spielt?

	Arabisch			Deutsch			Andere Sprache: Englisch		
Aktivitäten	0 Nie oder fast nie	1 Wenigstens einmal in der Woche	2 Jeden Tag	0 Nie oder fast nie	1 Wenigstens einmal in der Woche	2 Jeden Tag	0 Nie oder fast nie	1 Wenigstens einmal in der Woche	2 Jeden Tag
a. Lesen (Bücher, Magazine, Comics, Zeitungen, Zeitschriften)									
b. Fernsehen/Filme/Kino									
c. Geschichten erzählen/erzählt bekommen									
Gesamt									
Gesamt nach Sprache	/6			/6			/6		

5.3 Freunde der Familie, die oft zu Ihnen zu Besuch kommen benutzen:

	0 Nie	1 Selten/ ab und zu	2 Etwa die Hälfte der Zeit	3 Oft	4 Sehr oft/ immer
Arabisch					
Deutsch					
Andere:					

6. Informationen über die Eltern

6.1 Informationen über die Mutter

6.1.1 In welchem Land (+Ort) wurden Sie geboren? _____

6.1.2 Falls Sie zurzeit arbeiten, welche Sprache sprechen Sie im Beruf? _____

6.1.3 Ausbildung:

		Wie lange (Jahre)	Weitere Informationen
Grundschule	Ja / Nein		
Weiterführende Schule	Ja / Nein		
Universität	Ja / Nein		
Andere Berufsausbildung	Ja / Nein		

6.1.4 Wie gut schätzen Sie Ihre Kenntnisse in den folgenden Sprachen ein (betreffendes ankreuzen)?

	0 Gar nicht (höchstens ein paar Wörter)	1 Gering (Komme klar, aber habe Schwierigkeiten)	2 Mittel (Komme klar, ganz ok)	3 Gut (fühlt sich wohl)	4 Sehr gut (spricht fließend)
Arabisch					
Deutsch					
Andere Sprache					

6.2 Informationen über den Vater

6.2.1 In welchem Land (+Ort) wurden Sie geboren? _____

6.2.2 Falls Sie zurzeit arbeiten, welche Sprache sprechen Sie im Beruf? _____

6.2.3 Ausbildung:

		Wie lange (Jahre)	Weitere Informationen
Grundschule	Ja / Nein		
Weiterführende Schule	Ja / Nein		
Universität	Ja / Nein		
Andere Berufsausbildung	Ja / Nein		

6.2.4 Wie gut schätzen Sie Ihre Kenntnisse in den folgenden Sprachen ein (betreffendes ankreuzen)?

	0 Gar nicht (höchstens ein paar Wörter)	1 Gering (Komme klar, aber habe Schwierigkeiten)	2 Mittel (Komme klar, ganz ok)	3 Gut (fühlt sich wohl)	4 Sehr gut (spricht fließend)
Arabisch					
Deutsch					
Andere Sprache					

7. Besonderheiten/Schwierigkeiten

Bitte tragen Sie in jedes Kästchen JA (1 Punkt) oder NEIN (0 Punkte) ein:

	Geschwister	Mutter	Vater
Schwierigkeiten beim Lesen oder Schreiben			
Schwierigkeiten, andere zu verstehen, wenn sie in Ihrer Muttersprache sprechen			
Schwierigkeiten mit dem Sprechen (Aussprache, korrekte Sätze bilden, das richtige Wort zu finden, etc...) in der Muttersprache			
TOTAL			

8. Sonstige Bemerkungen: _____

IV BiLaD testing protocol (first page)

BiLaD: Protokollbogen für die Aufnahmen

Name des Kindes:

Kennung:

Geburtsdatum:

Alter 1. Sitzung:

Erstsprache(n):

Ort (Kita/Praxis/Schule):

Geplanter Ablauf:

Session 1		Session 2		Session 3		Session 0 Für Bi-TD & Bi-SLI	
Task	Time	Task	Time	Task	Time	Task	Time
LiSe-DaZ 1	7	Dog Story	5-10	NWRT	10-15	Bird Story	5-10
LiSe-DaZ 2	10	SRT	10				
Exhaustivity Part A	20	CPM	15	WWT	20	L1 Tests	
Card Sorting	5	PLAKSS	5	Exhaustivity B	10	L1 Tests	
Digit span	5	Hidden Mouse	5	Mr Peanut	10	Evtl. CPM	
LiSe-DaZ 3	4						
LiSe-DaZ 4	6						
Sum	50-60		50-55		50-55		

Tatsächlicher Ablauf:

Session 1		Session 2		Session 3		Session 0 (Bi-TD & Bi-SLI)	
Task	Time	Task	Time	Task	Time	Task	Time
Sum							

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich diese Dissertation selbstständig ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Alle den benutzten Quellen wörtlich oder sinngemäß entnommenen Stellen sind als solche einzeln kenntlich gemacht.

Diese Arbeit ist bislang keiner anderen Prüfungsbehörde vorgelegt worden und auch nicht veröffentlicht worden.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Lina Abed Ibrahim

A handwritten signature in black ink, consisting of stylized, cursive letters, positioned above a horizontal line.

Oldenburg, den 26.04.2026