Rima Haddad

Child bilingualism in Sweden and Lebanon

A study of Arabic-speaking 4-to-7-year-olds
Abstract


This dissertation investigates the vocabulary and narrative skills of 100 Arabic-Swedish-speaking children (aged 4–7 years) in Sweden cross-sectionally and the development of these skills (4 to 6) in a subgroup of 10 children longitudinally. Also, the vocabulary skills of 100 Arabic-speaking bilingual children (aged 4–7 years) in Lebanon are investigated cross-sectionally and compared to the Swedish cross-sectional study. Parental questionnaires were used to gather background information concerning language input and use inside and outside the home. The comprehension and production of vocabulary was assessed with the Cross-linguistic Lexical Task (CLT; Haman et al., 2015) and narrative macrostructure with the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2019). In Sweden, both Arabic and Swedish were investigated for vocabulary (language differences, age, socio-economic status (SES) and language input) and for narrative macrostructure (language differences, age and task effects). In Lebanon, Arabic vocabulary skills were explored in relation to age, SES and language input.

Sweden: For both vocabulary and narrative macrostructure, development with age was not only evident in Swedish, but also in Arabic. Children scoring high on Arabic vocabulary comprehension and production were older and had parents speaking with them mostly in Arabic. Joint book reading in Arabic boosted the children’s Arabic expressive vocabulary whereas being exposed predominantly to Swedish had a negative effect. For Swedish, high scoring children were older and had an early age of onset of Swedish. Children who were mostly exposed to Arabic scored lower on Swedish vocabulary. Surprisingly, SES (parental education) did not predict any of the vocabulary scores. In line with international studies, narrative macrostructure production scores were generally low at this age for both languages, even for the oldest children, whereas narrative comprehension was generally well developed, even for the youngest children. The longitudinal study largely confirmed the results obtained in the cross-sectional study.

Lebanon: Similarly to the Swedish sample, older children scored high on Arabic receptive and expressive vocabulary, children whose parents spoke with them mostly in Arabic scored high on expressive vocabulary, and no effects of SES were found. Compared to children in Sweden, children in Lebanon code-switched many more nouns.

Keywords: language development, bilingualism, Sweden, Lebanon, Swedish, Arabic, preschool-aged children, early school-aged children, vocabulary, narrative macrostructure, CLT, MAIN, language exposure, joint book reading, code-switching, SES

Rima Haddad, Department of Linguistics and Philology, Box 635, Uppsala University, SE-75126 Uppsala, Sweden.

© Rima Haddad 2022

ISSN 1652-1366
URN urn:nbn:se:uu:diva-489076 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-489076)
Acknowledgements

My years as a doctoral student have come to an end, and what a journey it has been! Writing this dissertation would have been much more difficult had I not been surrounded by so many extraordinary people.

First and foremost, I am deeply grateful to my supervisor Ute Bohnacker and her invaluable support. Ute’s door stood always open (welcoming so many long and stimulating academic and non-academic discussions) and her encouragement was much appreciated when days were tough. I could truly not have wished for a better supervisor! Thank you for sharing your wisdom and for reading and commenting on my text to the smallest detail.

I am grateful for the Swedish Research Council Vetenskapsrådet (VR 421-2013-1309, Bohnacker) and the Bank of Sweden Tercentenary Foundation Riksbankens Jubileumsfond (P19-0644:1) for funding the BiLI-TAS project. I am also grateful for numerous stipends for giving me the opportunity to attend conferences, winter schools and workshops (Anna-Maria Lundins stipendiefond, Jerringfonden, Stiftelsen Lundborgska Idofonden, Nybergs stipendiestiftelse) as well as to collect data in Lebanon (Carlo Landbergs stiftelse). Thank you Olof Gjerdman’s stipendiefond for supporting my extra time as a doctoral student.

Furthermore, I want to thank my second supervisor Anette Månsson for providing appreciated support on issues related to the Arabic language and its varieties. I would also like to express my gratitude to Elke Montanari, my opponent at the mock defence, for giving constructive feedback on how to improve my dissertation. Also, thank you Najwa Nasr, my teacher/supervisor at the BA and MA levels at the Lebanese University, for igniting in me a spark for linguistics and for encouraging me to pursue an academic path.

I would personally like to thank each individual child and her/his parents for taking part in this project. Without your contribution, this study would not have been possible. A heartfelt appreciation goes to everyone who helped recruit the children, including representatives of associations, congregations and schools, both in Sweden and in Lebanon. I genuinely appreciate your enthusiasm for the study and your support for me. Out of respect for the children’s anonymity, names of the organisations and their head leaders or group leaders will not be mentioned.

It was a real pleasure working with the BiLI-TAS team members starting with the head of the team (Ute Bohnacker), my fellow doctoral students (Josefin Lindgren, Buket Öztekin, Linnéa Öberg), researchers and research assistants (Sibylle Dillström, Karin Koltay, Frauke Jonsson, Zeinab Shareef, Amal Choumar, Alaa
Arifeh, Pascale Wehbe, Berrak Pinar Uluer) and the newest doctoral members in the team (Konstantina Daravigka and Pascale Wehbe). You have all contributed in crafting this dissertation while making the process of writing it very rewarding and fun. Thank you further to Josefin, Konstantina and Pascale for reading and commenting on my text at its final stages.

I would like to extend my sincere thanks to the many people who have contributed with their Arabic competence. You have enriched this project as well as my knowledge in the Arabic language and its vernaculars. A special šukran goes to Sarah Kohail, Zeinab Shareef, Lina Abed Ibrahim, Ziad Alqiss, Hana Berglund, Vima Haddad, Homam Alkordi and Nadim Chemaly. Also, šukran ġazilan Joseph Saouk and Ahmed Ruby for your feedback on my Arabic texts. A special merci ktir also goes to the great research ladies at the Lebanese University and the USJ in Lebanon who were always quick in replying supportively to my voice messages and emails.

To my former and current colleagues (and guest researchers who decorated the department along their way) at the Linguistics and Philology department, thank you all for making the everyday much more amusing! My list of colleagues who have inspired me throughout these doctoral years goes on and on. You know who you are and what good laughs we have had. Skål för alla goda minnen! It was also a pleasure working together with Karin Koltay, Miryam de Lhoneux and Klara Bertils leading the department choir LingSing. Thank you also to the dedicated and talented participants of the choir who put their trust in us. Together we have created fun and lasting memories for the whole department. Best choir ever?

As for life outside of academia, a heartfelt thank you goes out to my relatives and friends for checking up on me and encouraging me. A special gratitude goes to Christine Bou Rached, Chadia Haddad and Milma Rantamäki Danielsson whose friendship throughout the years has been a blessing. Thanks also to my friends in Uppsala (including members of Grekiska kulturhuset and Imm elNour) who made life in the city much more colourful.

I would like to thank my parents, Wadad and Ghassan, and siblings, Hala (Dr ḥalto) and Richard, for always supporting me, each in their own way. A special appreciation goes to teta Dedo (mom) who is the reason behind what I have accomplished in my life. She has always encouraged me to aim high and provided me with unconditional support to fulfil my dreams. Eliyas Lucas, Maria, Mikaela, Angelo Aram and Natalia, merci for constantly making my heart smile.

My greatest gratitude goes to my husband Nadim (my forever ‘José’) who joined me on this adventure in Uppsala. Thank you for encouraging me non-stop and standing firm by my side, especially during the most stressful periods. Nadim’s role was also key in the accomplishment of the Lebanese study. Nadim, without your support, completing this dissertation would not have been possible!

Sofia (ḥobbi), thank you for bringing so much joy into my life. Ya *juni!

Uppsala, November 2022, Rima Haddad
1. Introduction ......................................................................................... 13
   1.1 The study ............................................................................................ 16
   1.2 Arabic-speaking bilinguals in Sweden ............................................... 19
       1.2.1 The Arabic and Swedish languages ............................................ 22
       1.2.2 Language development of Arabic-Swedish-speaking children
            in Sweden ............................................................................................ 25
   1.3 Bilingualism in Lebanon .................................................................... 28
       1.3.1 Recent history of bilingualism and bilingual education in
            Lebanon ............................................................................................... 29
       1.3.2 Language development of bilingual children in Lebanon .......... 30
   1.4 Code-switching ................................................................................... 32

2. Aims and research questions ............................................................... 37

3. Previous research ................................................................................. 39
   3.1 Vocabulary ......................................................................................... 39
       3.1.1 Age.............................................................................................. 40
       3.1.2 Age of onset (AoO) .................................................................... 40
       3.1.3 Exposure ..................................................................................... 44
       3.1.4 Siblings ....................................................................................... 53
       3.1.5 Parent-child joint book reading .................................................. 57
       3.1.6 Mother tongue instruction (MTI) ................................................ 62
       3.1.7 Socio-economic status (SES) ...................................................... 63
   3.2 Narrative macrostructure .................................................................... 69
       3.2.1 Narrative comprehension with MAIN ........................................ 72
       3.2.2 Narrative production with MAIN ............................................... 77

4. Methods ............................................................................................... 84
   4.1 Recruitment ........................................................................................ 84
       4.1.1 Recruitment for the cross-sectional study in Sweden ................. 84
       4.1.2 Recruitment for the longitudinal study in Sweden ..................... 87
       4.1.3 Recruitment for the cross-sectional study in Lebanon ................ 89
       4.1.4 Summary..................................................................................... 89
   4.2 Materials ............................................................................................. 90
       4.2.1 Cross-linguistic Lexical Task (CLT)................................................. 90
       4.2.2 Multilingual Assessment Instrument for Narratives (MAIN).... 96
4.3 Procedure .......................................................................................... 100
4.3.1 The pilot study .......................................................................... 100
4.3.2 The cross-sectional study in Sweden ........................................ 101
4.3.3 The longitudinal study in Sweden ............................................ 103
4.3.4 The cross-sectional study in Lebanon ....................................... 104
4.3.5 Counterbalancing ...................................................................... 105
4.4 Transcription .................................................................................... 106
4.5 Scoring ............................................................................................. 110
4.5.1 Scoring of CLT ......................................................................... 111
4.5.2 Scoring of MAIN ...................................................................... 113
4.6 Ethics considerations ........................................................................ 118

5. Participants ........................................................................................ 121
5.1 Participants of the cross-sectional study in Sweden ......................... 121
  5.1.1 General information .................................................................. 121
  5.1.2 Age of onset and rated language proficiency ............................ 123
  5.1.3 Social background and language knowledge of the parents .... 125
  5.1.4 Language use in the home ...................................................... 127
  5.1.5 Home language use outside the home ..................................... 130
  5.1.6 Additional language input sources in the home ....................... 132
  5.1.7 Mother-tongue instruction ..................................................... 132
  5.1.8 Estimated daily language exposure ........................................ 133
  5.1.9 Attitudes towards language ................................................... 134
  5.1.10 Summary of the participants in the cross-sectional study in Sweden ........................................................................... 135
5.2 Participants of the longitudinal study in Sweden ............................. 137
  5.2.1 General information ............................................................... 137
  5.2.2 Language input and use in and outside the home ..................... 139
  5.2.3 Rated language proficiency .................................................... 141
  5.2.4 Language input at school and MTI ......................................... 143
  5.2.5 Estimated daily language exposure ........................................ 144
  5.2.6 Parents’ and teachers’ attitudes towards the home language .... 144
  5.2.7 Summary of the participants in the longitudinal study in Sweden ........................................................................... 144
5.3 Participants of the cross-sectional study in Lebanon ....................... 145
  5.3.1 General information ............................................................... 145
  5.3.2 Age of onset and rated language proficiency ............................ 146
  5.3.3 Social background and language knowledge of the parents .... 148
  5.3.4 Language use in the home ...................................................... 149
  5.3.5 Estimated daily language exposure ........................................ 152
  5.3.6 Attitudes towards language ................................................... 153
  5.3.7 Summary of the participants in the cross-sectional study in Lebanon ........................................................................... 154
6. Cross-sectional study in Sweden: Vocabulary

6.1 Research questions

6.1.1 Vocabulary scores and age
6.1.2 Vocabulary scores and background factors

6.2 Operationalization of variables and statistical analyses

6.2.1 Operationalization of variables
6.2.2 Statistical analyses

6.3 Results

6.3.1 Overview of the Arabic and Swedish vocabulary results
6.3.2 Vocabulary scores and age
6.3.3 Vocabulary scores and background factors

6.4 Discussion

6.4.1 Age
6.4.2 Age of onset
6.4.3 Socio-economic status (SES)
6.4.4 Input
6.4.5 Arabic varieties
6.4.6 Concluding remarks

7. Cross-sectional study in Sweden: Narrative macrostructure

7.1 Narrative comprehension

7.1.1 Research questions
7.1.2 Data and statistical analyses
7.1.3 Results

7.2 Narrative production

7.2.1 Research questions
7.2.2 Data and statistical analyses
7.2.3 Results

7.3 Discussion

7.3.1 Narrative comprehension
7.3.2 Narrative production
7.3.3 Concluding remarks

8. Longitudinal study in Sweden: Vocabulary and narrative macrostructure

8.1 Research questions
8.2 Background of the children
8.3 Results: Vocabulary

8.3.1 Overview of the Arabic and Swedish vocabulary results
8.3.2 Comparing the longitudinal study with the cross-sectional study
8.3.3 Development of vocabulary comprehension
8.3.4 Development of vocabulary production
11.2 Lebanon ........................................................................................................... 361
11.2.1 Characteristics of the Arabic-speaking bilinguals in Lebanon ......................... 362
11.2.2 Arabic vocabulary development with age ......................................................... 363
11.2.3 Code-switching ............................................................................................ 364
11.2.4 Influence of background factors on Arabic vocabulary development .................. 365
11.3 Methodological considerations and limitations .................................................. 368
11.3.1 Recruitment of participants ........................................................................... 368
11.3.2 Materials ..................................................................................................... 370
11.3.3 Questionnaire .............................................................................................. 371
11.3.4 Procedure .................................................................................................... 372
11.4 Future studies ................................................................................................. 372
11.5 Conclusions ................................................................................................. 374

12. (Summary in Arabic) ......................................................................................... 376

Appendix ............................................................................................................... 382
Appendix 1 (Chapter 4, Methods) .......................................................................... 382
A1.1 Counterbalancing tables .................................................................................... 382
A1.2 Parental education (SES) .................................................................................. 383
A1.3 Transcriptions (The cross-sectional study in Sweden: narrative macrostructure) ................................................................. 385
A1.4 MAIN Comprehension Questions ................................................................... 389
Appendix 2 (Chapter 7, Cross-sectional study in Sweden: Narrative macrostructure) ...................................................................................................................... 391
A2.1 Comprehension scores .................................................................................... 391
A2.2 Goals in narrative comprehension .................................................................... 391
A2.3 Production scores ............................................................................................ 392
Appendix 3 (Chapter 8, Longitudinal study in Sweden) ........................................... 394
A3.1 Macrostructural components ......................................................................... 394
A3.2 Story complexity .............................................................................................. 395
Appendix 4 (Chapter 9, Cross-sectional study in Lebanon: Vocabulary) ..................... 396
A4.1 Arabic vocabulary comprehension and production scores in Lebanon and Sweden ................................................................................................. 396
A4.2 Arabic vocabulary production and conceptual production scores in Lebanon and Sweden ................................................................................................. 397
A4.3 Age of onset (AoO) of Arabic and a second language ....................................... 397
A4.4 Children’s language use siblings ...................................................................... 398

Bibliography ........................................................................................................ 400
1. Introduction

In Sweden, bilingualism amongst children is steadily increasing. According to Statistics Sweden (2020), 38% of the children (age 0–18 years) residing in Sweden are considered to have a ‘foreign background’ (utländsk bakgrund in Swedish), i.e. either the child or both parents were born outside of Sweden. It can be assumed that most of these children, in addition to the majority language, Swedish, are exposed to another home language.\footnote{Note that children who have \textit{one} parent born abroad are not included in the ‘foreign background’ category (Statistics Sweden, 2020). Hence, the percentage of children exposed to another language in addition to Swedish is likely to be even higher.} Furthermore, 29% of school-age children (age 7–15 years, grades 1–9) are reported to have a home language other than, or in addition to, Swedish (National Agency for Education, 2021a). Of these minority home languages, Arabic is by far the most prevalent, where 25% of the children attending förskoleklass (a preparatory class between preschool and primary school) and 23% of the children in grades 1–9 are Arabic speakers. Due to recent large migration waves to Sweden from Arabic-speaking countries, Arabic is now the second largest home language after Swedish (National Agency for Education, 2021a).

On a global scale, bilingual individuals are believed to outnumber the monolinguals (e.g. de Bot & Kroll, 2002; Pearson, 2007; Grosjean, 2010). In fact, in many parts of the world, growing up as a bilingual is perceived to be the norm rather than the exception. Lebanon is one of these places where children are exposed to two (or more) languages from a very young age. Lebanon, unlike its neighbouring Arab countries, has a long and continuous history of bilingualism. The focus on bilingualism and the introduction to a second language usually begins before school age, at home or in preschool, and is later enforced at school.

Irrespective of country of residency or what the community’s policy on bilingualism is (Grosjean, 1982), bilingual children’s developmental trajectories in their two languages are influenced by similar types of factors. Some researchers classify these factors as child-internal and child-external. Child-internal factors include chronological age and age of onset, while child-external (or environmental) factors include those factors that determine the quantity and quality of the language input that the child receives (Paradis, 2011). In what follows, selected factors that have been found internationally
to affect the language-learning trajectory of bilingual children are briefly mentioned. For a more detailed discussion, see Chapter 3.

• Chronological age: As individuals get older, their cognitive abilities advance. These general cognitive abilities (e.g. better reasoning skills, increased working memory, richer long-time memory) may support language learning and use. As a consequence, children might need different amounts and types of language input at different ages (Rowe, 2012).

• Age of onset (AoO): Not all bilingual children acquire both of their languages at the same time. Some children are considered to be simultaneous bilinguals since they were exposed to two languages from birth (or from a very young age), while other children are considered to be sequential bilinguals since they were exposed to their second language (late AoO) some time after birth or during childhood. AoO may play a role in a bilingual child’s language development (e.g. Gagarina & Klassert, 2018; Gross, Buac & Kaushanskaya, 2014; Paradis, 2011). AoO is sometimes recalculated as length of exposure (LoE) by subtracting age of onset (usually in months) from the age of testing the child.

• Input quantity: Since a bilingual child’s waking time is divided between two languages, the quantity of language input is distributed between the languages. This has been shown to influence proficiency, especially with regard to vocabulary (e.g. Pearson, Fernández, Lewedeg & Oller, 1997; Thordardottir, 2011). Input quantity can be measured in several ways, for instance through proportion of exposure to a language (Thordardottir, 2011), parental or maternal language input (Sorenson Duncan & Paradis, 2020a) or language use at home (de Houwer, 2007; Gathercole, Thomas, Roberts, Hughes & Hughes, 2013). Without enough language input, children do not reach a comfort level that makes them willing to use the language (Pearson et al., 1997).

• Input quality: Input quality is related to the complexity, diversity and richness of language input. Not receiving quality input at a young age makes it more likely that children lag behind their peers in language skills in preschool or elementary school (Hart & Risley, 1995). Input quality may include the use of sophisticated language, rare words and decontextualized language (Rowe, 2012).

• Additional input opportunities: These features are a combination of input quality and quantity and include factors that have been shown to act as influential predictors for language development. Examples of additional language exposure opportunities include input from (older) siblings (Quiroz, Snow & Zhao, 2010; Sorenson Duncan & Paradis, 2020b), parent and child joint book reading (Sénéchal & LeFevre, 2002; Breitfeld, Potter &
Lew-Williams, 2021) and attending mother tongue instruction (Ganuza & Hedman, 2019).

- Socio-economic status (SES): SES is often said to be associated with input quality (Hart & Risley, 1995; Scheele, Leseman and Mayo, 2010). Many international studies have found that bilingual children from high-SES families often perform better than peers from low-SES families on vocabulary tests in the societal language (Buac, Gross & Kaushanskaya, 2014; Cobo-Lewis, Pearson, Eilers & Umbel, 2002a; Leseman, 2000), but that both groups perform alike in the home language (Buac et al., 2014; Prevoo et al., 2014). Alternatively, children from low-SES homes outperform children from high-SES homes in some aspect in the minority language (Cobo-Lewis, Pearson, Eilers & Umbel, 2002b).

- Language status: In many cases, bilingual children speak their mother tongue (which usually has a low status in society) merely at home and the majority language (which has a high status) in society. In Sweden, Swedish is the majority language. Children who live in such societies are believed to need more input in the minority language since attraction to the majority language (which is present everywhere around them in the environment) is very powerful (Pearson, 2007). However, in some countries, bilingual children live in a society where more than one language is accepted. In Lebanon, for example, Arabic as well as English and French enjoy high status and are widely spoken in society, and code-switching between these languages is embraced (Bacha & Bahous, 2011; Shaaban, 2017).

- Acquisition setting: In countries where the home language is not the majority language, parental input may be the only source of language input. This may lead to an unbalanced distribution of the language for home vs. school vocabulary (Montanari, Abel, Graßer & Tschudinovski, 2018). However, when a minority language has a relatively high number of speakers, children may be able to speak their L1 outside of home with other interlocutors as well. Although Arabic is a minority language in Sweden, Arabic-speaking children may find several opportunities to speak Arabic, for example with peers at (pre)school or at associations and congregations where Arabic is spoken in addition to the majority language.

- Language relatedness: The language combination of a bilingual child, especially when the languages are closely related, may facilitate the language acquisition process. Closely related languages that share cognates act as facilitators in learning vocabulary in both of the languages and especially boost the weaker one (Lindgren & Bohnacker, 2020).
1.1 The study

The present study investigates the language development of two Arabic-speaking bilingual groups: the first group is Arabic-speaking bilingual children who have Swedish as their second language and the second group is Arabic-speaking bilingual children who have English or French as a second language.

The first group includes Arabic-Swedish-speaking children residing in Sweden. The language development of this bilingual group is in focus in the present dissertation, in the sense that the children’s lexical skills and their narrative abilities are investigated in both of their languages. The Arabic-Swedish-speaking children are participants of the BiLI-TAS research project (Bohnacker, 2013) at Uppsala University. The BiLI-TAS project aims to explore the language development of Turkish-Swedish-speaking and Arabic-Swedish-speaking bilingual children with and without a developmental language disorder (DLD). For each of these language combinations in the BiLI-TAS project, a cross-sectional study, a longitudinal follow-up study and a clinical study have been carried out. The cross-sectional study includes approximately 100 typically developing (TD) children aged 4–7. The children’s lexical knowledge, narrative ability and phonological working memory are investigated in both of the children’s languages using assessment tools that are comparable across both languages. In the longitudinal follow-up study, approximately 10 4-year-old children from the cross-sectional study are tested again around two years later using the same testing materials. Finally, in the clinical study, children who have been diagnosed with DLD are assessed with the same testing materials as the children in the preceding two studies, thus allowing for a comparison of the language skills between children who have a DLD diagnosis and those who do not (and presumably have a typical language development). In all studies, the children are tested by trained native or near-native speakers of the children’s respective languages. Furthermore, background information concerning the children’s language development is gathered through an extensive parental questionnaire. The present study includes 100 Arabic-Swedish-speaking children from the cross-sectional study and 10 children from the longitudinal follow-up study in Sweden.

The second bilingual group participating in the present study comprises Arabic-speaking children residing in Lebanon. These children speak English

---

2 BiLI-TAS stands for ‘Bilingualism and Language Impairment, Turkish, Arabic, Swedish’ where Ute Bohnacker is the principal investigator. The research project was funded by the Swedish Research Council (Vetenskapsrådet, VR 421-2013-1309) 2014–2019 under the title of ‘Language impairment or typical language development? Developing methods for linguistic assessment of bilingual children in Sweden’. A continuation of the project is funded by the Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond P19-0644:1) under the title ‘Heritage language and Swedish language development from preschool to primary school.’

3 Typically developing refers to children who have not been diagnosed with Developmental Language Disorder (DLD, also known as Language Impairment, LI).
or French as a second language and are also aged 4–7. The data was collected by the author as a private initiative. Since language exposure is often believed to play a central role for the lexical development of bilingual children, studying the lexical development of Arabic-speaking bilingual children who have different language exposure possibilities might further highlight the importance of exposure. Hence, the purpose behind including the second bilingual group in the present study is to examine the effect that the linguistic environment and language exposure patterns have on two different groups of Arabic-speaking bilingual children who are the same age and are tested using the same tasks. Furthermore, the Lebanese Arabic variety belongs to the three most frequent Arabic varieties spoken in Sweden (see Section 1.2).

In general, little is known about the language development of these two Arabic-speaking groups. Although research has started to bloom within the past few years, more can be done. In Sweden, despite Arabic being a widely spoken language among minority groups, research on this population is still limited. Few studies have examined the lexical skills of Arabic-Swedish-speaking children, especially in relation to background factors. Furthermore, studies in Sweden regarding children’s narrative abilities are scarce (and mostly limited to work within the BiLI-TAS project). No extensive research on narrative skills has been done so far on Arabic-Swedish-speaking children in Sweden.

In Lebanon, research on children’s vocabulary has focused on development with age and only a limited number of studies have taken into consideration background factors that might influence vocabulary development.

Although the language exposure setting and the history of bilingualism in Sweden and Lebanon are different from each other (see Section 1.2 and Section 1.3), the language learning trajectories of bilingual children may be affected by child-internal and child-external factors in similar ways. By including Arabic-speaking children residing in Lebanon in the present study, the effect of certain background factors on the lexical development of two different types of bilingual language learning settings is better understood.

The language domains investigated in the present dissertation are vocabulary and narrative macrostructure. Vocabulary is a central language domain that allows individuals to understand and communicate. Vocabulary has been argued to be highly dependent on language input (e.g., de Houwer, 2007; Gathercole & Thomas, 2009; Gathercole et al., 2013; Hoff et al., 2012; Hoff et al., 2014; Pearson et al., 1997; Pearson, 2007; Thordardottir, 2011). For the assessment of vocabulary in the present study, the Cross-linguistic Lexical Task (CLT; Haman, Łuniewska & Pomiechowska, 2015) was used to examine

---

4 The idea evolved out of numerous discussions with parents of the children in the cross-sectional study in Sweden. Several parents expressed concern regarding their children’s knowledge in speaking their mother tongue since they did not live in their home country where the children have more opportunities to be exposed to Arabic.
the children’s lexical comprehension and lexical production. Measuring both vocabulary comprehension and vocabulary production was necessary since language exposure may influence comprehension and production differently (e.g. Thordardottir, 2011). Information regarding children’s language input was gathered (foremost) via parental questionnaires and analysed in relation to the children’s vocabulary scores.

As children become older, narrative competence becomes increasingly important for their social interactions, as well as for their educational success such as literacy development and reading comprehension (e.g. Dickinson & Tabors, 2001; Griffin, Hemphill, Camp & Wolf, 2004; Gutiérrez-Clellen, 2002). Compared to vocabulary, narrative competence has been argued not to be as much affected by input and language background, but more to be driven by general cognitive processes (i.e. the universality of narrative macrostructure, e.g. Gutiérrez-Clellen, Simon-Cereijido & Wagner, 2008; Iluz-Cohen & Walters, 2012; Pearson, 2002; Pearson & de Villiers, 2006; Uccelli & Páez, 2007; Paradis et al., 2011). For the investigation of the children’s narrative skills (narrative production and comprehension), the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina, Klop, Kunnari, Tantele, Välämaa, Bohnacker & Walters, 2019) was used. The present study focuses on the analysis of narrative macrostructure, i.e. the higher-order organization of narratives.

Both language tools (CLT and MAIN) were developed within the COST Action IS0804 research network (Armon-Lotem, de Jong & Meir, 2015) and were constructed to be comparable across the languages of bilingual children. Both tools are part of the LITMUS test battery (Language Impairment Testing in Multilingual Settings, see www.bi-sli.org and Armon-Lotem & Grohmann, 2021). Since Arabic and Swedish are typologically distant languages (see Section 1.2.1), analysing vocabulary and narrative abilities, instead of other language domains such as morphology or syntax, was judged to be more suitable for comparison. Both Arabic-speaking bilingual groups (in Sweden and in Lebanon) were assessed in Arabic with the CLT and MAIN assessment tools. However, for the cross-sectional study in Lebanon, only the vocabulary (CLT) results are reported. Furthermore, the children in Lebanon were tested in Arabic (and not also in their second language).

To sum up, the dissertation generally aims to provide a better understanding of Arabic-speaking bilingual children’s language development within the

---

5 A detailed description of the CLT and MAIN tools is found in the Methods chapter (Chapter 4).

6 The data collection in Lebanon took place during two visits to the country (see Section 4.1.3). Recruitment started already in Sweden to facilitate execution sur place. Since the data collection in Lebanon was based on a private initiative, it was not logistically possible (limited human resources and budget constraints) to train two experimenters (one English-speaking and another French-speaking) to assess the children in their second language. (In Sweden, the children were tested two times, one time in each language. This was not feasible in Lebanon.)
lexicon and narrative macrostructure. Specifically, a cross-sectional group consisting of Arabic-Swedish-speaking children ($N = 100$, age 4–7) in Sweden is assessed in both of the children’s languages for both their vocabulary and narrative skills. A subgroup of the children in the Swedish cross-sectional study are tested again two years later to study the development of vocabulary and narrative macrostructure in time. The dissertation further aims to understand the effect of background factors on the lexical development of the Arabic-speaking bilingual children in Sweden and in Lebanon.

This dissertation contains twelve chapters and is divided as follows. In Chapter 1, the rationale behind the study is described and the two Arabic-speaking bilingual groups and their languages are introduced. In Chapter 2, the aims and the research questions are specified. Chapter 3 gives a review of the literature related to (bilingual) children’s lexical knowledge and narrative skills. Chapter 4 describes the methods used in the cross-sectional study in Sweden, the longitudinal study in Sweden, and the cross-sectional study in Lebanon, when it comes to recruitment of participants (Section 4.1), description of the materials used (Section 4.2), procedures for data collection (Section 4.3), transcription of the data (Section 4.4), scoring of the data (Section 4.5), and ethics considerations throughout the study (Section 4.6). Chapter 5 characterizes the participant groups of all three studies. The following two chapters report the results for the cross-sectional study in Sweden regarding vocabulary (Chapter 6) and narrative macrostructure (Chapter 7). Chapter 8 covers the results of the longitudinal study in Sweden in terms of vocabulary development (Section 8.3) and narrative macrostructure (Section 8.4). In Chapter 9, the vocabulary results of the cross-sectional study in Lebanon are reported and compared with the vocabulary results of the cross-sectional study in Sweden. Chapter 10 provides a summary of the results while Chapter 11 includes a general discussion and conclusions. The dissertation ends with a summary in Arabic (Chapter 12).

After this general introduction, the remainder of this introductory chapter is divided into three parts. In Section 1.2, an overview of the Arabic-speaking population in Sweden is provided in addition to brief descriptions of the Arabic and the Swedish languages, and a review of previous studies on the language development of Arabic-Swedish-speaking children. Then, in Section 1.3, bilingualism in Lebanon is introduced, along with a depiction of Lebanon’s recent history of bilingualism. The section continues with an overview of previous studies on the language development of bilingual children in Lebanon. Finally, in Section 1.4, code-switching in bilinguals is discussed.

1.2 Arabic-speaking bilinguals in Sweden

This section contains a brief description of the Arabic-speaking population in Sweden. The information is based on an unpublished internal report by Ute
Bohnacker, the principal investigator of the BiLI-TAS project (Bohnacker, 2017). After this follows a brief typological description of the Arabic (Section 1.2.1.1) and Swedish (Section 1.2.1.2) languages and an overview of studies that have targeted the language development of Arabic-Swedish-speaking children (Section 1.2.2).

In Sweden, Arabic is the second-most frequent language (after Swedish) in terms of native speakers (National Agency for Education, 2021a). Although there are no official statistics concerning the numbers of speakers of a certain language in Sweden, this number may be inferred by, for example, finding out how many inhabitants were born in a certain country. According to Statistics Sweden, as of 31 December 2016, there were 362,597 residents in Sweden who were born in a country where Arabic is the main language (Statistics Sweden, 2017). These residents came predominately (86%) from Syria, Iraq and Lebanon. The high number of residents originating from these countries is the result of three major migration waves to Sweden.

Sweden’s history of mass migration from countries where Arabic is a main language dates back to the 1980s. The first migration wave was from Lebanon as a consequence of the Lebanese civil war (1975–1990). The second wave of migration came from Iraq in the early 1990s and 2000s as a consequence of the Gulf war and the invasion of Iraq. The third wave, which is the latest and largest migration wave, came from Syria after the outbreak of the civil war in 2011. In 2016, the number of residents in Sweden born in Lebanon made up 26,906 individuals, those born in Iraq 135,129, and those born in Syria 149,418 (Statistics Sweden, 2017). Thus, the two main reasons for emigration to Sweden from Arabic-speaking countries have been seeking asylum (fleeing war, conflict, persecution and hardship) and family reunification (permit for refugees to unite and live with close family members).

Another way to estimate the number of Arabic speakers in Sweden is to see how many individuals were born in Sweden who have one or both parents from an Arabic-speaking country. In 2016, 94,712 individuals were registered as born in Sweden having both parents originating from the same Arabic country, 55,315 individuals were born in Sweden having parents originating from two different Arabic countries, and 30,012 individuals were born in Sweden with one parent from an Arabic country and one Sweden-born parent (Bohnacker, 2017, based on Statistics Sweden, 2017).

As a consequence of the three major migration waves from Arabic-speaking countries, the Arabic varieties mostly spoken in Sweden are the Levantine (approx. 48%) and the Iraqi (approx. 37%) varieties. However, individuals who were born in countries where Arabic is the official language are not guaranteed to have Arabic as their mother tongue. Aramaic/Syriac, Armenian and Kurdish are additional languages spoken by minority groups in the Arab world, especially in the Middle East and Iraq. However, due to the ‘Arabization programs’ that emerged after the colonization period and whose aims include the protection and promotion of the Arabic language (Bentahila,
one can assume that most of the migrants from Arabic countries are speakers of Arabic, even if some of them are bilinguals having a minority language as their mother tongue.

Adding up all the above-mentioned numbers (number of residents born in an Arabic-speaking country and number of Sweden-born residents whose parents were born in Arabic-speaking countries), the number of Arabic speakers in Sweden is likely to exceed 400,000. According to Statistics Sweden (2017), these individuals live predominately in Sweden’s three largest cities and their suburbs, namely Stockholm, Göteborg and Malmö. Additionally, Södertälje, a city southwest of Stockholm, hosts the highest numbers of Syria-born individuals and the fourth highest number of Iraqi-born individuals in Sweden.

Another way to estimate the Arabic-speaking population in Sweden is via statistics concerning the attendance of *modermålsstöd* ‘mother tongue support’ in nursery school and *modersmålsundervisning* ‘mother tongue instruction’ (MTI) in *förskoleklass* ‘pre-school class’ and school. For simplicity, these two terms will be referred to as MTI henceforth. Sweden has offered MTI since the 1970’s for children who have a mother tongue other than Swedish. MTI is provided in extra classes of typically 40–60 minutes/week organized by the municipality. The main aim behind MTI is the development of oral and literary proficiency in the child’s minority language. In the academic year of 2016/2017, the number of students nationwide who were eligible for MTI in Arabic was 64,261.7 Out of all the children who were entitled to attend MTI in their respective languages (275, 329), 23% concerned Arabic, making it the most spoken mother tongue amongst children in Sweden and it has been so for the last ten years. Additionally, MTI in Arabic had the highest attendance rate with 43% in pre-school class and 66% in school (National Agency for Education, 2017).8

In short, the number of Arabic-speakers in Sweden is unknown but is estimated to exceed 400,000, with origins predominately in Syria, Iraq and Lebanon. The Arabic language is the most common language among children eligible for MTI as well as the one with the highest attendance rate amongst pupils.

7 Recent numbers from the National Agency for Education (2021a) indicate that during the school year of 2020/2021, Arabic was the most spoken mother tongue (after Swedish). Around 80,500 elementary school children were eligible for MTI classes in Arabic (25.7% of all children who have the right to MTI in Sweden), and around 58,700 (i.e. 72.9 %) attended.
8 No numbers are available regarding children who are eligible for *modermålsstöd* ‘mother tongue support’ in preschool (age 1–6 years). Many municipalities, including Stockholm and Uppsala, have reduced their support for the teaching of the mother tongue before children start pre-school class (age 6 years).
1.2.1 The Arabic and Swedish languages

In this section, an introductory description is provided about the language characteristics of Arabic (Section 1.2.1.1), followed by Swedish (Section 1.2.1.2).

1.2.1.1 Arabic

Arabic is a Semitic language belonging to the Afro-Asiatic language family. In 2013, it was estimated to be spoken as a native language by 300 million speakers and as a second language (L2) by an additional 60 million speakers (Owens, 2013, p. 2). Arabic is spoken in large parts of the world, particularly in the Middle East and Northern Africa, but also by the diaspora who migrated mainly to Europe and America.

The spoken everyday Arabic language, ʿāmiyya, comprises many different vernaculars (spoken varieties) that differ from the standard written Arabic variety, fuṣḥa or Modern Standard Arabic (MSA). MSA has a prestigious position and it is reserved as the language of religion, culture and education whereas colloquial Arabic is used for everyday communication. MSA and spoken Arabic differ from each other in the fields of vocabulary, phonology, syntax, and grammar (Abu-Rabia, 2000). Ferguson (1959) used the term diglossia to highlight the dichotomy between high (MSA) and low (colloquial) varieties of Arabic used for different functions and situations. Several scholars have since then published extensively on the nature of diglossia in Arabic, including Badawi (1973) who instead emphasized the continuum from the colloquial local variety to a regional variety towards a more formal MSA. For an overview, see Versteegh (2014, Chapter 12). Inhabitants of villages that are geographically and historically close will face no difficulty in understanding each other’s vernaculars, however, this ease in communication may disappear the farther away one moves geographically. Traditionally, the Arabic varieties are classified into five groups: varieties of the Arabian peninsula, Mesopotamian varieties, Syro-Lebanese varieties, Egyptian varieties and Maghreb varieties (Versteegh, 2014).

9 Versteegh (2014) compares the difference between MSA and colloquial Arabic varieties to that between Latin and French (p. 241). He also provides examples of how the different Arabic varieties differ from MSA (pp. 134–137) and from each other (Chapter 11).

10 See Thomure, Tamim and Griffiths (n.d.) for an overview of studies regarding the teaching and learning of the Arabic language (MSA) in native Arabic-speaking school children.

11 As previously mentioned, Arabic speakers in Sweden mainly originate from the Levant and Iraq. Arabic varieties spoken in Lebanon, Syria, Jordan and Palestine belong to the ‘Syro-Lebanese varieties’ group. Although dialects spoken within these regions are usually classified into three groups (Lebanese/Central Syrian, North Syrian and Palestinian/Jordanian), there is no clear-cut distinction between the groups. The varieties spoken in Iraq (often classified as gilit and qəltu based on the varieties found in Baghdad) belong to a different group, namely the ‘Mesopotamian varieties’ group (see Versteegh, 2014, pp. 197–205). The term used by Versteegh was ‘dialect’ and not ‘variety’. 
In general, colloquial Arabic is used for oral communication throughout society and is learnt from parents, family members and the community, while MSA is mainly used for reading and writing and is taught formally at school. Hence, no child or adult has MSA as their mother tongue; however, many children may be exposed to MSA from an early age via, for example, religious services and the media (such as TV or the internet). Similarly, Arabic speakers may also be exposed to other Arabic varieties than their own through contact with peers (especially while living as migrants) and through the globalization of the media. Interlocutors who speak different Arabic varieties might temporarily need to eliminate their local vernacular language features and replace them with more regional or MSA ones in order to facilitate communication.

Arabic is written and read from right to left using the Arabic alphabet. The alphabet has 28 letters that are connected in cursive style where words are separated through spaces. The letters change in their written form based on their position in the word (initial, medial, final or isolated form). At times of script unavailability or technical difficulties, colloquial Arabic is often written using the Latin alphabet, for example, in mobile text messages.

Words in Arabic usually consist of a consonantal root (of 3 or 4 consonants, also called radicals) which carries the lexical meaning, combined with a vowel pattern of short and/or long vowels spread between the consonants. Arabic is a highly inflected language where grammatical information (e.g. tense/aspect, person, gender, number) in the verbal, nominal and adjective domains is mainly determined via affixes. Words in phrases show clear concord/agreement between nouns and their dependents (usually adjectives, nouns or pronouns) and between verbs and their subjects. Whilst Arabic has a rich pronominal system, subject pronouns are often omitted as Arabic is a pro-drop language. Arabic has two genders (feminine and masculine) and three numbers (singular, dual and plural). Furthermore, Arabic has three cases (nominative, genitive, accusative) which are marked by case endings in MSA but generally not as much in the spoken Arabic varieties. There are two verbal aspectual forms, imperfective and perfective. The future tense is formed by adding a prefix or a future auxiliary to an imperfect verb. Other tenses also exist and are also usually formed through compound tenses involving auxiliary verbs and particles. Noun and adjective definiteness is marked by an enclitic prefix article (al-) while indefiniteness is not marked. The Arabic language has prepositions. The linking of two nouns in a construct phrase (iḍāfa) is often used to show relations of possession and related semantic functions (such as identity, partitive, and purpose). There are two types of clauses, nominal clauses (containing no verb and no copula in the present tense, but

---

12 The *hamza* [‘], usually pronounced as a glottal stop, may be considered as a letter on its own, thus making the count 29 letters.

13 In colloquial Arabic, adjectives are usually not inflected for dual but rather for plural (e.g. *baytēn kbār* in Lebanese Arabic instead of MSA dual *baytān kabīrān* ‘two big houses’).
including a copula in the past tense) and verbal clauses. The basic word order of verbal clauses in MSA and colloquial Arabic is Verb-Subject-Object, although Subject-Verb-Object is usually more common in spoken Arabic.

1.2.1.2 Swedish
Swedish is a North-Germanic language belonging to the Indo-European language family. There are an estimated 10 million speakers of Swedish (Swedish Institute, 2021). Swedish is primarily spoken in Sweden as well as Finland and to a small degree by the Swedish diaspora, most of whom reside in the Nordic countries.

Unlike Arabic, there is no diglossia in Swedish, and spoken and written Swedish are much closer to each other than MSA and colloquial Arabic. Phonological alterations, colloquial speech or slang expressions do occur in oral speech but are usually not written in formal writing. For example, in colloquial Swedish, one could say Dom ser dom instead of the written form De ser dem ‘they see them’.

Swedish is written and read from left to right using the Latin alphabet. The Swedish alphabet consists of 29 letters which include the 26 letters of the English alphabet with the addition of three diacritic vowel letters (å, ä, ö).

Words are made of roots (or stems) to which affixes can be added, and compounding, i.e. the combination of several word stems is widespread. Swedish has two genders (neutrum, en-words and utrum, ett-words) and two numbers (singular and plural) that affect the grammatical form of nouns and adjectives. Definiteness is expressed by the addition of a suffix to the noun and indefiniteness is expressed by the insertion of an article (en, ett) before the noun. Swedish is sometimes described as having two cases, or no case at all. The nominative is never marked, and the genitive (either considered a case or a clitic particle) is marked by an –s suffix.

Swedish is not a pro-drop language, the presence of both subject and verb is obligatory (i.e. there is a subject-verb constraint). Unlike Arabic, Swedish always uses overt pronouns and copulas. Verbs, regardless of tense, stay unchanged irrespective of gender, person or number. Swedish verbs mainly have three inflectional forms, infinitive, past and supine. The remaining verb forms may be derived from these three inflections and tense formation and may or may not involve an auxiliary. Verb conjugations are classified into four major divisions (three weak verb and one strong verb class). Swedish has many prepositions where the most easily identifiable prepositional relations are those related to time, place and manner (Holmes and Hinchliffe, 2013).

A frequent word order in Swedish declarative clauses is Subject-Verb-Object. Other word orders are also found but the verb has always the second position (V2-rule). An exception is yes/no questions where the verb takes the initial position and is directly followed by the subject (e.g. Gillar du musik? = like you music ‘Do you like music?’).
To sum up, Arabic and Swedish are not related languages since they originate from different families and until recently have had very little contact. Very few cognates exist between the two languages, hence, knowing one language will not help in the acquisition of the other language in terms of vocabulary. The languages are read and written from different directions, using different scripts. As for grammatical aspects, both Arabic and Swedish nouns accept affixes, show definiteness, use pronouns though to varying degrees; however, they differ in gender, number and infinitive marking. Arabic, unlike Swedish, is a pro-drop language and information regarding number and gender is shown through the affixes. Swedish has a fixed word order, something which is more flexible in spoken Arabic.

1.2.2 Language development of Arabic-Swedish-speaking children in Sweden

Studies on the vocabulary skills of Arabic-Swedish-speaking children are relatively few and include unpublished undergraduate and graduate theses, a few research articles and recent publications by the BiLI-TAS team. Below is an overview of selected studies that have investigated the lexical knowledge of Arabic-Swedish-speaking children in Sweden. Studies on Arabic-Swedish-speaking children’s narrative skills in relation to narrative macrostructure do not exist yet.

In their BA thesis, Mikoczy and Nyman (2008) compared the Swedish (L2) receptive vocabulary (PPVT-III, Dunn & Dunn, 1997) and vocabulary organization (Kent & Rosanoff, 1910) of 49 Arabic-Swedish-speaking bilingual children (age around 10–11) with the receptive vocabulary of monolingual Swedish-speaking children and other Swedish-speaking bilingual children in Sweden from previous studies. Results showed that the Arabic-Swedish-speaking children scored significantly lower on both vocabulary measures than children in previous studies.

Similarly, in another BA thesis, Alkass Yousef and Bergström (2011) explored the receptive and expressive vocabulary of 16 children (age 6;2–7;0), 9 of whom were Swedish-speaking monolinguals and 7 were Arabic-Swedish-speaking bilinguals. For both languages, receptive vocabulary was measured with the PPVT-III (unpublished translations of Dunn & Dunn, 1997) and expressive vocabulary was measured with Ordracet (Eklund, 1996). Results showed that the monolinguals scored significantly higher on Swedish receptive vocabulary than what the Arabic-Swedish-speaking bilinguals scored on Arabic receptive vocabulary. Furthermore, monolingual children scored significantly higher in both Swedish and Arabic receptive vocabulary than what the Arabic-Swedish-speaking bilinguals scored in both Swedish and Arabic receptive vocabulary.

---

14 The Arabic version of the PPVT-III was translated by Eva Kristina Salameh and the Arabic version of Ordracet was translated by Sabina Alkass Yousef. The bilingual children were not tested for Swedish receptive vocabulary.
significantly higher in Swedish expressive vocabulary than the bilingual children, but not so when conceptual scoring was applied, meaning that when the Arabic and Swedish answers were added up, there was no significant difference between the two groups. The bilingual children scored significantly higher in Swedish than in Arabic in expressive vocabulary.

Salameh (2011) examined the lexical knowledge of 49 Arabic-Swedish-speaking bilinguals (age 10–11) where one group (N = 33) had received monolingual Swedish schooling and a smaller group of children (N = 16) had received bilingual Arabic/Swedish schooling. The children were tested for their receptive vocabulary (PPVT-III, unpublished translations of Dunn & Dunn, 1997) in both of their languages. Salameh found that the children who had attended monolingual instruction scored higher on Swedish receptive vocabulary than the children who had attended bilingual instruction but the difference between the two groups was not significant. Similarly, the children who had attended bilingual instruction scored higher on Arabic receptive vocabulary than the children who attended monolingual instruction, but the difference between the two groups was not significant either.

In his master’s thesis, Ridha (2015) investigated lexical, morphological and syntactic aspects of Arabic stories told by 12 Iraqi Arabic-Swedish-speaking children (age 5–7) using the MAIN (Gagarina et al., 2012) as a narrative elicitation tool. Ridha found that crosslinguistic influence was most frequent on the lexical level and that the children code-switched not only from Iraqi Arabic to the majority language Swedish while telling the MAIN story, but also occasionally to MSA and other Levantine Arabic varieties.

In the BiLI-TAS project, the language development of bilingual groups of children in Sweden has been investigated. One hundred Arabic-Swedish-speaking children have been tested for their lexical and narrative knowledge. Information regarding the children’s lexical knowledge was gathered with the CLT (Haman et al., 2015) and background information via a parental questionnaire. See Chapter 4 for a detailed description of the test materials and the questionnaire. So far, few aspects regarding the lexical knowledge of the Arabic-Swedish-speaking children have been analysed.

In her doctoral dissertation, Öberg (2020) examined the relationship between lexical knowledge of 99 Arabic-Swedish-speaking children (age 4;0–7;11) and three background factors (daily exposure, length of exposure to Swedish and SES). Öberg found that, for Swedish, percentage of daily exposure (as a linear variable) significantly predicted Swedish expressive (but not receptive) vocabulary scores, while for Arabic, percentage of daily language exposure was a significant predictor of both receptive and expressive vocabulary. As for SES (parental education, as a linear variable), Öberg found no

---

15 Conceptual scoring takes into consideration the meaning of the response regardless of the language in which it is produced, i.e. the number of concepts lexicalized by the child (Pearson et al., 1993).
correlation between SES and vocabulary scores in either language. Length of exposure to Swedish, however, was a significant predictor of both receptive and expressive Swedish vocabulary.

These results were expanded on in Bohnacker, Haddad and Öberg (2021). Moreover, Bohnacker et al. (2021) illustrated the importance of individual differences by providing case studies of four children and discussed how background factors (e.g. different history of language exposure, current language use) might contribute to children’s vocabulary proficiency.

Bohnacker, Haddad, Lindgren, Öberg, Öztekin (2020) analysed Swedish vocabulary development in Swedish-speaking monolingual children \( (N = 72) \), Arabic-Swedish-speaking bilingual children \( (N = 76) \), and Turkish-Swedish-speaking bilingual children \( (N = 74) \) aged between 4;0 and 6;11. Bohnacker and colleagues found a clear age effect for all groups in both receptive and expressive vocabulary, however there was also a significant difference between the monolingual and bilingual children. For receptive vocabulary, both bilingual groups showed a stronger age effect than the monolinguals who performed at ceiling or near ceiling already at age 4. For expressive vocabulary, the Turkish-Swedish-speaking bilingual children showed a stronger age development than the Arabic-Swedish-speaking bilinguals whose group mean score was lower. In their discussion, Bohnacker et al. mention a difference between the two bilingual groups’ age of onset of Swedish as a possible explanation for why the Turkish-Swedish-speaking group performed better than the Arabic-Swedish-speaking bilinguals in Swedish expressive vocabulary.

In another comparative study between 100 Arabic-Swedish-speaking children and 102 Turkish-Swedish-speaking children (age 4;0–8;1) from the BiLI-TAS project, Bohnacker, Haddad, Lindgren, Öberg, Öztekin (2021) explore how the children’s vocabularies in both languages are affected by various background factors. For instance, Bohnacker et al. found that, for the majority language, length of exposure to Swedish affected both groups’ Swedish vocabulary. For the minority language, children of parents who spoke only/mostly the mother tongue with them, scored significantly higher than children whose parents did not. Furthermore, Arabic-Swedish-speaking children who attended MTI had larger Arabic receptive and expressive vocabularies than children who did not. However, the same effect for MTI attendance was not found for the Turkish-Swedish-speaking children.

In her master’s thesis, Wehbe (2020) investigated the relation between expressive vocabulary of nouns and narrative character introduction in 29 Arabic-Swedish-speaking children (age 5;0–6;11) from the BiLI-TAS project, 20 of whom had typical language development (TD) and 9 had a DLD-diagnosis. Narrative samples were gathered with MAIN (Gagarina et al., 2019). For both bilingual groups, Wehbe found a strong correlation between Arabic expressive vocabulary and character introduction in Arabic, but not between Swedish expressive vocabulary and character introduction in Swedish.
To sum up, several studies have investigated the lexical knowledge of Arabic-Swedish-speaking children in Sweden in the past, yet many of them include a small sample of children and none of them investigate the influence of age and background factors on the children’s lexical development. Recent publications by the BiLI-TAS research group are based on a large sample of children (included in the current study) and investigate the effect of selected background factors in relation to vocabulary scores. However, since language input occurs from several sources, further background factors (such as shared book reading and language input from siblings) still need to be investigated in relation to vocabulary. No study has investigated the production and comprehension of narrative macrostructure skills in Arabic-Swedish-speaking children yet.

In the current study, the lexical and narrative skills of 100 Arabic-Swedish-speaking children are investigated in both of their languages. The children’s vocabulary skills are explored in relation to a range of background factors. Furthermore, the effect of background factors on the L1 (Arabic) of the Arabic-Swedish-speaking children in Sweden is compared to the effect that background factors have on Arabic-speaking bilinguals in Lebanon, a country where bilingualism is wide spread and accepted.

In the next section, bilingualism in Lebanon and its history will be discussed, followed by an overview of studies in direct relation to the lexical proficiency of Arabic-speaking bilingual children in Lebanon.

1.3 Bilingualism in Lebanon

Similarly to other Arab countries, Lebanon’s national official language is Modern Standard Arabic (MSA), however the language that people actually speak is colloquial Lebanese Arabic. MSA (fuṣḥa) is learned formally at school and is used mostly for written, religious, educational and media purposes. The Lebanese colloquial Arabic (ʿāmiyya or dāriǧ) is used everyday in informal domains.

Lebanon has a history of being multicultural and multilingual and this is reflected in daily language use as well as the languages used in the educational systems (schools and universities). In addition to Arabic, French and English are often spoken in Lebanon and hold high status in society. Hence, children in Lebanon grow up as bilinguals (or multilinguals) since they are exposed to two (or more) languages from a very young age, often before entering preschool (at age 3 or 4).16

---

16 Armenian, a minority language in Lebanon, is also used as a language of education in some private schools.
1.3.1 Recent history of bilingualism and bilingual education in Lebanon

Bilingual education in schools has been present in Lebanon since the nineteenth century, when Lebanon was under Ottoman rule (1516–1918), through the settlement of European (predominately French) and American Christian missionaries (Shaaban & Ghaith, 1999). Lebanese Christians established contact with European churches and started to adopt French as a cultural language among the educated (Versteegh, 2014). After World War I, Lebanon underwent a French mandate rule (1920–1943) where both Arabic and French became the official languages of the country and both languages were compulsorily taught in private schools (Jarrar, Mikati & Massialas, 1988, cited in Bacha and Bahous, 2011). After Lebanon’s independence (1943), the French language lost its status as the co-official national language; however, it continued to have a central role in education and in public life since it was by now the language for the French-educated ruling elite (Zakharia, 2017). A few years later (1946), English was introduced as an official alternative second ‘foreign’ language alongside French to be taught in schools (Shaaban & Ghaith, 1999).

Ever since the Lebanese civil war (1975–1990), the English language has expanded in popularity, at the expense of French, which has lost its dominant position in the country (Versteegh, 2014). As of 1994, all schools in Lebanon (both public and private) have to teach Arabic, and a first ‘foreign’ language (English or French) which is the main medium of instruction, as well as a second ‘foreign’ language (English or French) which starts at early primary level but is taught only around three lessons per week. The Lebanese curriculum (issued in 1997) gives equal weight to the native language, Arabic, and the first ‘foreign’ language (English or French) in terms of numbers of teaching hours throughout a child’s whole school years, i.e. from preschool to the final year of secondary school (Shaaban & Ghaith, 1999). Today, only few school subjects have to be taught in Modern Standard Arabic, namely history, geography [and civics] and Arabic language and literature, while the remaining subjects are usually taught in either English or French (Thonhauser, 2001). Thus, around 77% of school instruction is in English or French and only 23% is in standard Arabic (Hoyek, 2004).

---

17 Lebanon’s neighbouring country, Syria, took a very different path after the French mandate period, essentially that towards monolingualism, by erasing all French language traces and making Arabic the exclusive medium for instruction in the public educational systems and in the media (Bentahila, 2011; Versteegh, 2014).

18 Private schools with a Muslim profile differ in their hour distribution for teaching the Arabic language, giving it more time throughout the week since Arabic is the language of the Koran (Bahous, Bacha & Nabhani, 2011).

19 These numbers may vary according to school type (e.g. public or private), education cycle (e.g. elementary, intermediate, secondary) and education type (formal, non-formal or vocational education) (Hoyek, 2004).
Schools in Lebanon are of two types, public ones which are run by the government and make up about 40% of the schools in Lebanon, and private ones which make up around 60% of the schools (Hoyek, 2004; World Bank, 2021). The fee-paying private schools usually enroll children from high socio-economic families whereas the public schools are usually attended by children from low socio-economic families. Private schools are often believed to offer a higher quality of education than public schools (Shaaban, 2017; Bahous & Nabhani, 2008).

All the children participating in the Lebanese part of the current study attended private schools (with either English or French as a first ‘foreign’ language, i.e. main language of instruction). See Chapter 5 for more information about the participants in the Lebanese cross-sectional study.

1.3.2 Language development of bilingual children in Lebanon

Until recently, only few studies have investigated the typical language development of Lebanese Arabic-speaking children. Those concerning vocabulary and language proficiency are summarized below.

Zablit’s (2008) master’s thesis examined the estimated expressive vocabulary of 210 toddlers (a younger group, age 17–19 months and an older group, age 26–28 months) residing in Beirut and its suburbs. The children were monolingual Arabic speakers (N = 72), monolingual French speakers (N = 74) and bilingual Arabic-French speakers (N = 64).21 The parents were asked to fill in the MacArthur-Bates Communicative Development Inventory (CDI, Fenson et al., 1993) in all of the children’s respective languages. Results showed that the older children had significantly higher estimated vocabularies than the younger children, regardless of whether the children were monolingual or bilingual. When the bilinguals’ vocabulary was scored for both languages combined (i.e. conceptual scoring), their vocabulary size became comparable to their French monolingual peers’, whereas the lexicon of the Arabic-speaking monolingual children was significantly smaller than that of the other groups. A complication in Zablit’s data is that the socio-economic background of the groups differed. Although no statistical analyses were made, Zablit argues that her findings can be explained by the Arabic-speaking children’s

---

20 Bahous et al. (2011) mention another division, namely an approximate 50% split between public and private schools. The data collection in Lebanon for the present dissertation occurred prior to the ongoing Lebanese economic crisis (from 2019). According to a recent World Bank report (2021), the crisis in Lebanon has led to an upheaval in school attendance. During the school year of 2020–2021, around 737,000 children, i.e. more than one third of the school-age population (36% of all children aged between 3–18 years) residing in Lebanon, did not attend school, furthermore, about 55,000 children (11%) moved from a private to a public school.

21 Zablit (2008) differentiated between bilinguals and monolinguals according to which exposure parents reported for their children. Children who had received 80% or more exposure in one language were considered monolinguals in that language.
low-SES (parental education), language input, and the difference in morphological richness between the two languages. For all groups, parents estimated the children to know more nouns than verbs. However, Zablit found that the Arabic-speaking monolinguals scored higher on verbs than the other two groups (Zablit, 2008).

Daccache, Messarra and dos Santos (2020) examined the estimated expressive and receptive vocabulary of 112 Arabic-speaking bilingual toddlers (age 8–16 months, mean age 11.7 months) in Lebanon with adapted versions of the CDI-I (Fenson et al., 1993) parental checklists in Arabic, French and English. The parents filled in the checklists for all three languages. Results showed that even at this young age, there was a correlation between the children’s age and their estimated receptive and productive vocabulary in the total scores of all languages combined and for each language measured alone. The authors explain this by the toddler’s constant exposure to at least two languages daily, regardless of whether the child had Arabic as her/his L1 or any of the other two ‘foreign’ languages (Daccache et al., 2020).

Zebib, Henry, Messarra, Kouba Hreich and Khomsi (2019) carried out a large-scale age norming study using a Lebanese adaptation of the French screening test Evaluation du Langage Oral (ELO; Khomsi, 2001) named Evaluation du Langage Oral chez l’enfant Libanais (ELO-L; Zebib, Henry, Khomsi, Messarra & Kouba-Hreich, 2017). The tool was normed on 1,718 children (age 3–8 years) residing in Beirut, northern Lebanon and southern Lebanon attending both public and private schools. The test is composed of five subtests including tasks targeting receptive vocabulary and conceptual expressive vocabulary. For both vocabulary measures, a linear increase in performance with respect to the children’s age was found (Zebib et al., 2019).

Kanj and El-Hassan (2021) examined the expressive vocabulary of 74 Lebanese Arabic-speaking children (age 3–9; mean age 6;6) using the Lebanese Picture Naming Test, a picture naming tool (of semantically diverse nouns) especially constructed for the study. Children were asked to name 157 coloured pictures in any language they knew (Arabic, English or French). The children were divided into two SES groups based on whether they were sampled from private schools ($N = 47$ children sampled from 3 private schools) or public schools ($N = 27$ children sampled from 2 public schools) in Beirut. Results showed that age was a significant predictor of expressive vocabulary. Furthermore, the children’s SES (school type) was also a predictor where children who were enrolled in private schools scored significantly higher than children who attended public schools (Kanj & El-Hassan, 2021).

In a study that used the CLT to test the vocabulary skills of bilingual children, Khoury Aouad Saliby, dos Santos, Kouba Hreich, and Messarra (2017) investigated (Lebanese) Arabic comprehension and production of nouns and verbs of 32 typically developing (TD) Arabic-English/French bilingual children (age 5;7–6;9) and 10 DLD-diagnosed (Developmental Language Disorder) Arabic-English/French bilingual children (age 5;9–7;10) residing in
Beirut, northern Lebanon and southern Lebanon. Although the testing session was in Lebanese Arabic, both TD and DLD-diagnosed groups named many of their expressive vocabulary items in their L2 English or French. The authors found that during the Arabic verb production subtask, both TD and DLD-diagnosed groups answered significantly more often in Arabic than in their L2, whereas on the Arabic noun production subtask, only the TD children provided more answers in Arabic while the DLD-diagnosed children answered almost as often in their L2.

In sum, many of the studies conducted on the vocabulary development of Lebanese Arabic-speaking children concern young toddlers or are studies of children who have a large age range or include relatively few participants. Many of the studies merely adopt conceptual scoring (and thus accept answers in English and/or French when Arabic is tested, and vice versa). Although a few studies have taken into consideration certain background factors, none have done so extensively, for example, by analysing the effects of a child’s daily language exposure, both parents’ language input and shared-book reading on child vocabulary. The current study investigates the effect of background factors on bilingual children’s Arabic vocabulary in Lebanon. These results are also compared to the effects background factors have on Arabic-speaking bilingual children in Sweden, where Arabic is a minority language.

1.4 Code-switching

As previously mentioned (Section 1.3.2), in some studies carried out in a bilingual setting in Lebanon, when a child was asked to name an object in Arabic, s/he sometimes answered in English or French. This could be seen as an example of code-switching as the experimenter asked in Arabic but the child answered in English/French. It is well known that the extent to which code-switching is used depends on the interlocutors and the setting (Grosjean, 1982). In more monolingual settings there is less code-switching, whereas in settings that are more bilingual, code-switching is more common. Code-switching is widespread among bilingual children and virtually all children code-switch at some time (Paradis, Genesee & Crago, 2011, p.89). Bilinguals, as opposed to monolinguals, must choose which language to speak and decide whether or not to mix between their two languages based on the discourse situation and their interlocutor (Paradis, 2007).

A distinction is sometimes made in the literature (e.g. Ritchie & Bhatia, 2013) between the terms code-switching and code-mixing where the former implies a shift in language is made intersententially (i.e. across utterance/sentence boundaries) and the latter implies a shift intrasententially (i.e. within
the same utterance/sentence). In this dissertation, the term code-switching is used as an umbrella term for both code-switching and code-mixing.

Ritchie and Bhatia (2013) summarized the reasons why bilinguals code-switch and classified them into four factors (p. 378–388). The first factor is the interlocutors’ roles and the dynamic of their relationship. The second factor is situational (formality and setting), meaning that the choice of language changes according to discourse topic and language distribution (for example, shifting audience from private to public). Furthermore, social variables (such as group membership in religion, gender or age) can influence the pattern of code-switching both qualitatively and quantitatively. The third factor is message-intrinsic (linguistic considerations), such as mentioning a direct quotation, paraphrasing, hedging or interjections, for example, to emphasize what is said or to quote someone for a more authentic narration. Finally, factors to do with language attitude, language dominance and linguistic security play a role in a bilingual person’s language choice (Ritchie & Bhatia, 2013). Regardless of the reason why bilingual adults and children code-switch, code-switching is usually systematic and conforms to the grammatical constraints of the languages in question (Paradis et al., 2011, p. 103).

For example, below is a short segment from an Arabic narrative transcription (MAIN1, Cat) of a 5-year-old child (BiAra5-09) who participated in the cross-sectional study in Sweden. The code-switched words are in Swedish and marked with [@s].

(1) w baʿden (.) &el [//] fi waḥdi bisse w waḥdi ṣabi och [@s] wahdi, hon fi waḥdi bisse och [@s] waḥdi ṣabi, så [@s] elṣabi ska [@s] aḥud eltabi min elmay <och [@s] &el> [/] och [@s] elbisse ska [@s] akul elfiskar[@s] ...

‘and then (.) &el [//] there is one cat and one boy and [@s] one, here there is one cat and [@s] one boy, so[@s] the boy will [@s] take the ball from the water and [@s] the cat will [@s] eat the fish [@s] …’

This particular boy code-switches a number of function words (conjugation, auxiliaries) into Swedish, as well as a noun (fiskar ‘fish’). Other children may restrict their code-switching to nouns.

---

22 In the narratives by the Arabic-Swedish-speaking children, intra-utterance code-switching (within the same utterance) was more common than inter-utterance (between the utterances). See (i) for an example of intra-utterance code-switching by one 6-year-old child (BiAra6-03) when telling MAIN2 (Baby Birds) in Arabic. Swedish code-switches are marked by [@s].

(i) ba den iŋa &wakil [//] akil (.) <en (.) svans>@s] tabaʾ elkat[@s] &s &tt, baʿden sawala jaga[@s]

‘then came &wakil [//] food (.) <a (.) tail>@s] for the cat[@s] &s &tt, then he made chase[@s]’

---

23 Throughout the dissertation, examples of the children’s responses are provided. Swedish text is presented in standard Swedish orthography whereas Arabic text is written based on the transcription alphabet of the Encyclopaedia of Arabic Language and Linguistics (Eid et al., 2011).
There are several hypotheses why bilingual children code-switch (summarized by Paradis et al., 2011; pp. 92–103). One explanation for code-switching is the *Gap-Filling Hypothesis*, according to which a child code-switches more frequently when s/he is speaking in her/his non-dominant language and/or when the child does not know the appropriate word in the target language (or when an equivalent translation of the word does not exist). The above mentioned example illustrates the gap-filling hypothesis where the child did not know the Arabic equivalent of the words and decided to code-switch to Swedish. Other examples can be found in the cross-sectional study in Lebanon where many children did not know the Arabic label for the animals that were part of the vocabulary (CLT) production task. To illustrate, the target word *ḍifdaʿa* ‘frog’ was only named correctly in Arabic by 14 children out of a total of 100 children. Six children answered that they didn’t know the Arabic label and did not give a translated equivalent, 2 children provided an incorrect answer (misinterpreting the visuals of the image). However, seventy-eight (78 out of 100) children (from all age groups, 4–7) gave a code-switched correct answer in English (*frog*) or French (*grenouille*) instead of the target word in Arabic. None of these children were able to name the animal in Arabic when the experimenter asked the question again emphasizing that the child should provide the word in Arabic (rather than in another language).

Another reason why children code-switch may be related to *experiential factors* that result from interaction with the world and cause an uneven balance in a bilingual child’s lexicon. For instance, bilingual children who learn about the solar system at school in the majority language might have fewer opportunities to hear and learn the translation equivalent of the terminologies in their native language.

Finally, code-switching might also be the *social norm* (Grosjean, 1982) of the child’s family, community or society. Some of bilingual children’s communities are tolerant and accepting of code-switching whereas others are not. Children who are being brought up in bilingual families or communities where code-switching is the norm may have difficulty keeping their languages separate and it may take some time for them to accommodate to a monolingual setting (Paradis et al., 2011).

An example of a code-switching tolerating society is found in Malta where children grow up with two majority high-status languages, Maltese and English. Maltese adults often code-switch between the two languages when addressing children, for example by inserting English lexical items in Maltese syntactic frames (Gatt, Attard, Luniewska, & Haman, 2017). In a study that investigated the comprehension and production of nouns and verbs (using the CLT) of 5-year-old Maltese-English-speaking bilingual children (*N* = 56),

---

24 Note that although the children were tested in a ‘monolingual setting’ in each of their testing sessions (Maltese and English), they were tested by the same experimenter with a short break of 10 minutes in between. This is unlike the testing procedure used in the current dissertation.
Gatt et al. (2017) found that when conceptual scoring was taken into consideration in vocabulary production, children who were dominant in one language code-switched correctly (i.e. provided a correct conceptual answer but not in the target language) more when they were tested in their non-dominant language, however, more so in Maltese and less in the English CLT tasks. Gatt and colleagues concluded that mixing between the two languages appeared to be the outcome of lexical gaps in the tested language, hence triggering a retrieval of relevant labels in the other language (Gatt et al., 2017).

Similarly to the situation in Malta, children residing in Lebanon grow up bilingual (speaking two or even more languages) from an early age. Daily conversation is conducted in colloquial Lebanese Arabic but school education is mostly in English or French (Thonhauser, 2001). Many vocabulary assessment studies of Lebanese children take into account all of the children’s languages using ‘conceptual scoring’ (recall Section 1.3.2). In general, in Lebanese society, code-switching is seen as a natural and widespread phenomenon (Bacha & Bahous, 2011; Shaaban, 2017) that might even occur unconsciously (Bahous, Baroud Nabhani, Bacha, 2014). However, the bilingualism situation is at times believed to come at the expense of Lebanon’s national language, Arabic (Bahous, Bacha & Nabhani, 2011; Bacha & Bahous, 2011). Thonhauser (2001) stated that many Lebanese identify colloquial spoken Arabic as their mother tongue, but find it difficult to express themselves in written Arabic (MSA) since it is considered to be a distant language that is learned as a foreign language at school. Shaaban and Ghaith (1999) mention concerns about ‘semilingual’ speakers (as opposed to being multilingual) where individuals in Lebanon have ‘no real proficiency in any of the languages’. Other researchers have a more positive attitude towards code-switching saying that it is ‘the norm rather than the exception [… and that it] is one of the most distinctive features of the Lebanese culture’ (Grosjean, 1982), as well as that code-switching adds ‘richness to the language experience’ (Bahous, Baroud Nabhani & Bacha, 2014).

However, code-switching is not unique for those who benefit from speaking both of their languages in society, but as previously mentioned, a common feature amongst bilinguals. Even when the experimenter aims for a monolingual setting for the testing session, bilingual children may answer in the non-target language. Furthermore, the extent of code-switching may differ between children, between the language features and between language status (majority vs. minority language). For example, in Sweden, Sender and Svensson (2018) examined the non-target responses given by 52 Turkish-Swedish-

---

where there was interval of a mean of 10.7 days between the two testing sessions, each held by a different experimenter. See Chapter 4.
speaking children (age 4, \( N = 26 \); age 7, \( N = 26 \)) on the Swedish CLT production task.\(^{25}\) Code-switching instances, i.e. providing an answer in the non-target language (Turkish), made up 3.4% of the non-target responses. The majority of these came from the 4-year-olds, where nouns were code-switched 3 times more than verbs. Similarly, Ridha (2015), who investigated cross-linguistic influences in the Arabic narratives of Arabic-Swedish-speaking children (MAIN, Gagarina et al., 2012), noticed that Arabic-Swedish-speaking children code-switched many more nouns than verbs. Wehbe (2020) who studied the character introduction skills of 29 Arabic-Swedish-speaking bilingual children (with and without a DLD-diagnosis), also using MAIN (Gagarina et al., 2019) found that children code-switched more from Arabic to Swedish than the other way around, despite the monolingual setting of the testing situation.

In sum, code-switching is a phenomenon that normally occurs for bilingual children. How much a child code-switches might be related to several factors but code-switching from the minority language to the majority language as well as code-switching nouns more than verbs, seem the most common. In the present study, code-switching will be mainly studied in relation to children’s expressive vocabulary when analysing similarities and differences between the Arabic-speaking children in Sweden and those in Lebanon (Section 9.3.2.3).

\(^{25}\) The children are part of a larger Turkish-Swedish-speaking sample in the BiLI-TAS project that focuses on the Turkish-speaking children. See Öztekin (2019).
2. Aims and research questions

In this chapter, the main three aims and three research questions of the present study are presented.

The first aim is to explore the vocabulary and narrative abilities of Arabic-Swedish-speaking children age 4 to 7 residing in Sweden. A cross-sectional sample of 100 (typically developing) children are tested in both of their languages (Arabic, the home minority language, and Swedish, the majority language) with assessment tools that target the children’s lexical knowledge and narrative skills.

To better understand the children’s lexical knowledge, parallel and comparable tasks that measure noun and verb comprehension and production are used to investigate how vocabulary develops with age (in the cross-sectional sample), how the children perform in both languages, and whether they perform differently in comprehension and production. An additional interest is to understand to what extent certain background factors may be connected to vocabulary growth. Via information gathered through parental questionnaires on language use, language input and socio-economic status (SES), the children’s performance on the vocabulary tasks is analysed in relation to various background factors.

To explore the development of narrative abilities, the 100 children are asked to tell two oral stories in each of their languages, Arabic and Swedish. The production of narratives is elicited by parallel picture-based tasks that are comparable across the languages. The comprehension of narratives is tested via inferential comprehension questions. The stories told, as well as the children’s answers to comprehension questions, are analysed with a focus on macrostructure (i.e. the overarching structure of story events).

The second aim is to analyse how vocabulary and narrative abilities develop over time. A sub-group of ten 4-year-old children from the cross-sectional sample in Sweden is tested again after two years, at age 6, with the same tasks as in the first study. Potential changes in the children’s language input patterns and background factors are observed via a new parental questionnaire and a parental interview. Hence, effects of language input change are explored in relation to individual children. The development over time in these children is compared to the apparent development with age observed in the cross-sectional sample.
The third aim is to investigate how background factors affect the home language (L1, Arabic) development of Arabic-speaking children in two different types of bilingual settings. The effect on vocabulary performance in the cross-sectional sample in Sweden (where Arabic is a minority language) is compared to that in the cross-sectional sample in Lebanon (where Arabic is a majority language).

Thus, the three research questions (RQ) of the present study are:

- **RQ1:** What are the vocabulary and narrative abilities of Arabic-Swedish-speaking children (age 4–7 years) in both of the children’s languages, and which background factors measurably affect these abilities?
- **RQ2:** How do the vocabulary and narrative abilities of Arabic-Swedish-speaking children develop from age 4 to age 6 in the children’s two languages?
- **RQ3:** How do background factors affect the Arabic (L1) vocabulary development of Arabic-speaking bilingual children in Lebanon in comparison to the Arabic-Swedish-speaking children in Sweden?
This chapter is divided into two sections, each providing a review of the background literature on children’s language development. The first section focuses on vocabulary development (3.1) and the second on narrative macrostructure (3.2). Studies on bilingual children whose age ranges are of particular interest for the current study (4–7-years) are in focus, however, relevant studies on monolingual children or children who have a different age range are also mentioned.

3.1 Vocabulary

The trajectory of vocabulary learning starts from early childhood and continues throughout life into old age. Vocabulary is needed to comprehend and convey meaning as well as to communicate effectively in different contexts and about different topics. Vocabulary knowledge affects children’s abilities to understand words and use them properly during speaking and listening, and eventually during reading and writing at school age (Sinatra, Zygouris-Coe & Dasinger, 2012). Children understand more words than what they can produce; unsurprisingly, comprehension scores on receptive vocabulary tests are usually higher than production scores on expressive vocabulary tests.

Vocabulary development in bilingual children does not necessarily resemble that of monolingual children since certain background factors might impact their vocabulary development in each language differently. In this section, findings from earlier studies on various background factors (child-internal factors and child-external factors related to the environment) that have been shown to play a role for children’s vocabulary development and that are relevant for the current study are summarized.

First, language development through age will be briefly discussed (Section 3.1.1), followed by the influence of age of onset (AoO) on vocabulary development (Section 3.1.2). Then, environment-related factors that may affect language development are discussed, starting with language exposure (Section 3.1.3), language use with siblings (Section 3.1.4), parent-and-child joint book reading (Section 3.1.5), MTI attendance (Section 3.1.6) and SES (Section 3.1.7).
3.1.1 Age

As shown by numerous studies, vocabulary growth with age in bilingual children at preschool and school age occurs robustly in the majority language, yet often to a lesser extent in the minority language, despite substantial input, (e.g. Bialystok, Luk, Peets, & Yang, 2010; Cobo-Lewis et al., 2002a, 2002b; Dijkstra, Kuiken, Jorna, & Klinkenberg, 2016; Gagarina et al., 2017; Gauza & Hedman, 2019; Gathercole & Thomas, 2009; Gathercole et al., 2013; Hoff, Core, Place, Rumiche, Señor & Parra, 2012; Leseman, 2000; Lindgren, 2018; Lindgren & Bohnacker, 2020; Montanari, Akıncı & Abel, 2019; Zablit, 2008; Öztekin, 2019).

Similarly, for monolingual children, age has been found to be a significant predictor of vocabulary development (e.g. Bialystok et al., 2010; Cobo-Lewis et al., 2002a, Hoff et al., 2012, Lindgren, 2018). For instance, the Cross-linguistic Lexical Task (CLT) utilized in the current study was administered by Haman et al. (2017) to 639 monolingual children of 17 different languages (age 3;0–6;11) to test whether vocabulary scores correlate with the children’s age (in months). Results showed that when all language samples were considered (i.e. as one whole sample), there was a significant positive correlation between age and CLT vocabulary scores. When language samples were tested individually, 11 out of the 17 languages correlated with age, Swedish being one of them. The lacking effect of age development on vocabulary in the remaining languages was explained by small sample size, narrow age range or a combination of both factors. Hence, Haman et al. (2017) concluded that as monolingual children grow older through preschool to school age, they can be expected to increase their performance on the CLT.

Discussed below are a number of studies that have examined children’s lexical development in relation to background factors.

3.1.2 Age of onset (AoO)

The effect of age of onset (AoO) for a bilingual child’s two languages is a topic much explored. Many researchers have drawn a line of distinction between simultaneous bilinguals (acquiring both languages from birth) and sequential bilinguals (acquiring one language from birth and another language after some time) at the age of three (e.g. McLaughlin, 1978; de Houwer, 1995; Ruiz-Felter et al., 2016; Thordardottir 2019) and found contradictory results as to whether simultaneous bilinguals have an advantage over sequential bilinguals. Other studies that have used alternative cut-offs (e.g. Unsworth, 2016; Gross, Buac & Kaushanskaya, 2014; Gagarina & Klassert, 2018) also

---

26 The 639 children came from 15 different countries and spoke 17 different languages, namely Afrikaans, Catalan, English (British), English (South African), Finnish, German, Hebrew, isiXhosa, Italian, Lithuanian, Luxembourgish, Norwegian, Polish, Serbian, Slovak, Swedish, and Turkish. Arabic (Lebanese) was not part of this study.
found contradictory results. (For an extensive review on studies on simultaneous and sequential bilinguals, see Paradis, 2007). Finally, still other researchers claim that AoO by itself does not tell anything per se, but rather age of onset (or length of exposure to a language) mediated by other background factors. Presented below are studies that have studied the effect of age of onset on vocabulary.

Thordardottir (2011) examined the receptive and expressive vocabulary of both languages of English-French-speaking bilingual children (mean age 58.3 months) in Québec province (Canada), where both English and French are widely spoken in the community and both are majority languages of high status. All children had received regular input in both languages before the age of 3 (simultaneous bilinguals). The children’s receptive vocabulary was measured with the Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) in English and Échelle de vocabulaire en images Peabody (EVIP; Dunn, Theriault-Whalen & Dunn, 1993) in (Canadian) French. Expressive vocabulary was measured with a subtest from the Clinical Evaluation of Language Fundamentals (CELF-P, 2nd edition; Wiig, Secord & Semel, 2004) in English and the ‘expression-vocabulaire’ subtest N-EEL (Chevrie-Muller & Plaza, 2001) in French. When dividing the sample further into children with early onset of both languages, before age 6 months, and children with a later bilingual exposure, after 20 months, and while controlling for language input, Thordardottir found no difference in vocabulary scores in any language or vocabulary measure between the two groups.27

In another study in Canada, Paradis (2011) investigated whether length of exposure (in months) to English (L2) affected the English receptive vocabulary (PPVT-III, Dunn & Dunn, 1997) of 169 English-speaking bilingual children (age 4;10–7;0, with diverse L1s, range of exposure to English 3–62 months). In a regression model that included several child-internal and child-external factors as predictors for L2 vocabulary, length of exposure (in months) to English was shown to be a significant predictor. Paradis, however, continued on to say that the richness of the English environment (exposure quality) had a more significant impact on L2 vocabulary than length of exposure (exposure quantity), but that both were relevant to the determining rate of a child’s L2 acquisition.

In the US, Gross, Buac and Kaushanskaya (2014) examined the receptive and expressive vocabulary of simultaneous (N = 39) and sequential (N = 19) Spanish-English-speaking bilingual children (age 5–7). Unusually, the cut-off line between simultaneous and sequential bilinguals was whether the child was able to produce a 2-word sequence before or after the age of 3. Results

27 Thordardottir (2011) considers the children in her sample as simultaneous bilinguals and adds that this particular division for AoO (before age 6 months; after age 20 months) was mainly determined by the sample’s age distribution.
showed that for both receptive and expressive vocabulary in L2 (English), the simultaneous bilinguals significantly outperformed the sequential bilinguals (Gross et al., 2014).

Unsworth (2016) studied whether the age of onset of Dutch (L2) for 87 English-Dutch-speaking children, residing in the Netherlands, had any effect on the children’s Dutch receptive vocabulary (as measured with the PPVT-III-NL, Dunn, Dunn & Schlichting, 2005). Results showed that children who had an early AoO to Dutch (after age 1 but before 4, $N = 44$) did not score significantly higher than children who had a late AoO (at or after age 4 but before age 8, $N = 43$). Unsworth concluded that AoO was not a significant predictor of receptive vocabulary in the children’s L2 but rather that language input quantity was (Unsworth, 2016).

Gagarina, Posse, Gey, Golcher and Topaj (2017) examined the development of lexical abilities in German (L2) in relation to the AoO of 39 Russian-German-speaking children and 55 Turkish-German-speaking children (all age range 26–47 months, mean age = 38.5). Gagarina et al. found that early AoO of L2 negatively impacted the expressive vocabulary of the respective L1 for both bilingual groups.

Similarly, another large-scale study in Germany explored whether early or late AoO had an effect on the children’s L1 rather than their L2. Gagarina and Klassert (2018) examined the effect of AoO of German (L2) on the Russian (L1) receptive and expressive vocabulary of 213 Russian-German-speaking children (age 26–98 months, mean age = 52.8). Vocabulary was measured via the Russian Language Proficiency Test SPRUK (Gagarina et al., 2010). AoO was determined according to whether the children started to receive regular exposure to German before 18 months ($N = 31$), between 18 months and 3;05 years ($N = 90$) or between 3;06–5;05 years ($N = 25$). Correlation analyses showed that AoO of L2 correlated negatively with the children’s receptive and expressive vocabulary skills in their L1. However, when AoO was put in a regression model to examine the combined effect of additional background factors (age, gender, use of L1 in the family), AoO was only a significant predictor of expressive vocabulary and not for receptive vocabulary in L1. In other words, the later the child came in contact with German (L2), and while taking into consideration other background factors, the better the child performed in Russian (L1) expressive vocabulary (Gagarina & Klassert, 2018).

---

28 Receptive vocabulary was measured via the PPVT-III (Form B, Dunn & Dunn, 1997) for English and TVIP (Dunn, Padilla, Lugo & Dunn, 1986) for Spanish. Expressive vocabulary was measured via the Picture Vocabulary subtest of the Woodcock-Johnson III Test of Achievement (Form A, Woodcock, McGrew, and Mather, 2001) for English and Batería III Woodcock-Muñoz Pruebas de aprovechamiento (Muñoz-Sandoval, Woodcock, McGrew, and Mather, 2005) for Spanish.

29 The age of testing was controlled for: the early AoO children were tested at a mean age of 7;3 and the late AoO were tested at mean age of 9;1. Age of testing showed no significant effect on vocabulary results.
This is in contrast to an earlier finding by Gagarina, Armon-Lotem, Altman, Burstein-Feldman, Klassert, Topaj, Golcher and Walters (2014) where length of exposure to L2 (and the amount of L2 language spoken at home by the parents) had no negative influence on the Russian-German-speaking and Russian-Hebrew-speaking bilingual children’s L1 expressive vocabulary. Expressive vocabulary was measured with a German noun-verb naming task (Kauschke, 2007) adapted to Russian and Hebrew (i.e. not the same vocabulary test material as that used by Gagarina and Klassert, 2018).

In another study by Thordardottir (2019) in Québec, the receptive and expressive vocabulary of both languages of English-French-speaking bilingual children (in grades 1 and 3) was compared to that of monolingual French-speaking peers. The bilingual children had either an exposure to French before the age of three (\(N = 62\)) or after (\(N = 38\)). French receptive vocabulary was measured with the EVIP (Dunn & Thériault-Whalen, 1993) and expressive vocabulary with CELF-4 Canadian (Semel, Wiig, Secord, Boulianne, & Labelle, 2009). English receptive vocabulary was measured with the PPVT-III (Dunn & Dunn, 1997) and the expressive vocabulary with CELF-4 (Semel, Wiig, & Secord, 2003). Thordardottir concluded that AoO was not a significant predictor of French (L2) vocabulary, whereas cumulative amount of exposure to French was. Furthermore, both bilingual groups lagged behind the monolinguals in receptive vocabulary.

Öberg (2020), who studied a large sub-sample of the children participating in the current study, investigated the effect of Length of Exposure (LoE) of Swedish on the Swedish receptive and expressive vocabulary of 99 Arabic-Swedish-speaking children (by subtracting AoO from each child’s age in months). Öberg found that LoE was a significant predictor of both receptive and expressive vocabulary scores in Swedish, even when LoE was examined in regression analyses along with age, daily language exposure and SES (Öberg, 2020). Öberg did not examine the effect of Swedish LoE or AoO on the children’s L1 (Arabic) vocabulary.

In short, there seems to be no definite agreement on the cut-off point for early vs. late age of onset and whether this division has any a significant effect on vocabulary development. Studies that have used length of exposure (as a continuous variable) have also found conflicting findings. Note that in most of the above mentioned studies, the bilingual children lived in societies where the home language was a minority language, often not spoken in the broader society but mostly only at home. An exception was Thordardottir (2011 & 2019) who studied the effect of age of onset on the vocabulary of bilingual children in Québec, a province in Canada where both English and French are widely spoken. The situation in Lebanon is quite similar to that in Québec, where colloquial Arabic in addition to English and French holds high status in society (see Section 1.3.1), however, no studies have been found to have studied the effect of AoO on the lexical knowledge of Lebanese children. Finally,
the majority of the above reviewed studies have examined whether the exposure of one language has an effect on that particular language, yet have not tackled whether AoO has any significant effect on the child’s other language. In the current study, the effect of AoO will be investigated in both languages of the bilingual children in the Swedish cross-sectional study and merely in Arabic in the Lebanese cross-sectional study. Furthermore, the effect of AoO will be studied in relation to other language acquisition factors, such as language exposure, which will be discussed next.

3.1.3 Exposure

In addition to age and age of onset, vocabulary development has often been studied in relation to language exposure. As has been proven in numerous studies, vocabulary is a domain that is strongly affected by the proportion of exposure. When children receive more input in one language, their vocabulary grows and their vocabulary scores on assessment tests are expected to increase (e.g. Hoff et al., 2012; Pearson et al., 1997; Sorenson Duncan and Paradis, 2020b; Thordardottir, 2011; Unsworth, 2016). This relationship between input and vocabulary is described as a self-reinforcing cycle where more input promotes language proficiency resulting in more language use which in its turn invites more input (Pearson, 2007).

However, studies differ at times in their results regarding the effect of exposure, depending on which vocabulary domain is measured (receptive or expressive), which testing materials are used, how old the children are, and how the quantity and/or quality of language exposure are measured. For example, some studies measure language exposure (quantity) by the child’s total amount of exposure throughout the day, while others measure it by language use in the family at home. The choice of method for measuring language use (or test material) depends on the age of the children studied. Young toddlers interact daily with older individuals in their family who take care of them; hence, the family becomes the primary socializing and input agent for their development of bilingualism (de Houwer, 2009). Older children who attend (pre)school interact with more interlocutors throughout their day (multiple sources of language input for both L1 and L2). The vocabulary of young toddlers is usually measured with parental checklists in addition to parental interviews/questionnaires. Occasionally, some studies make use of recorded video or audio samples of parent-child interactions to examine language exposure quality. The vocabulary of older children is usually tested with picture identification or picture naming tasks for receptive and expressive vocabulary, respectively. Information on language input is typically gathered via parental reports (e.g. questionnaires) since this method of gathering information is believed to be cheap, fast and reliable (Paradis, 2017).
Below is an overview of selected studies that have investigated how children’s different vocabulary domains are influenced by various aspects of language exposure. Studies on bilingual and monolingual children from different age ranges and using different eliciting methods are summarized below, to give a broad idea of the large field of research on how different exposure patterns affect children’s language(s) and vocabulary modalities.

Pearson, Fernández, Lewedeg and Oller (1997) examined the relationship between language exposure and the receptive and expressive vocabulary of 25 Spanish-English-speaking bilinguals (age 8–30 months) in the US. The parents were asked to fill in a CDI checklist in both of the child’s languages (English: MacArthur Communicative Development Inventory, CDI, 1989; Spanish: Jackson-Maldonado & Bates, 1988) and provide a language measure estimate (via a questionnaire) of the child’s quantity of language input. Results (reported only for the children’s L1 Spanish) showed that the more time a child heard Spanish, the more Spanish words the child knew. Through longitudinal observations, the authors highlighted the effect of language environment changes on input and vocabulary acquisition. Additionally, Pearson and colleagues drew the conclusion that children who heard one of their languages less than 20% of the time were generally not willing to produce utterances in that language (Pearson et al., 1997).

In a large-scale study by Cobo-Lewis, Pearson, Eilers and Umbel (2002a, 2002b), 952 elementary-school students in the US (332 kindergarteners, 306 second-graders and 314 fifth-graders) were tested for their oral language, reading and writing skills. The majority of the children (N = 704) were Spanish-English-speaking bilinguals whereas the remaining children (N = 248) were English-speaking monolinguals. The children’s vocabulary knowledge was investigated with four different oral lexical tests in both languages.³⁰ Age (as measured via the children’s grade: kindergarten, grade 2 and grade 5) showed a clear effect on the majority language English, but not as much for the minority language Spanish. Language spoken at home was significantly associated with the bilingual children’s vocabulary scores where children who came from families who spoke both English and Spanish at home scored significantly higher in English than children who spoke only Spanish at home. This advantage, however, decreased as children grew older (as tested in the 5th grade). Conversely, language spoken at home was also associated with the bilingual children’s Spanish scores where children who came from homes where only Spanish was spoken outperformed children who came from English-Spanish-speaking households (Cobo-Lewis et al., 2002a, 2002b).

De Houwer (2007) studied the effect of different parental input patterns on 1899 Dutch-speaking bilingual children (age 6–10 years, with different L1s)

³⁰ The children were tested in English using the Woodcock Language Proficiency Battery (Woodcock, 1991) and the PPVT (Dunn & Dunn, 1981) and in Spanish with Spanish versions of the tests (Dunn et al., 1986; Woodcock & Muñoz-Sandoval, 1995).
in Flanders, Belgium. The parents were asked to fill in a questionnaire about the languages used at home and an estimation of their child’s L1 proficiency. De Houwer concluded that language use in the home was strongly related to the estimated language proficiency of the children in their L1. All children would eventually learn the majority language (L2), but the parent’s language use at home would determine whether the child learned the L1 or not. De Houwer also found that the children’s language use of Dutch at home differed from that of their parents’, but did not differ from the language use of their siblings (de Houwer, 2007).

Bialystok, Luk, Peets and Yang (2010) studied the English receptive vocabulary of 772 monolingual children in comparison to that of 966 bilingual children who had English as their L2 (and spoke diverse L1s). The children (age 3–10 years) were sampled from several earlier studies (conducted by the first author in Canada) and all tested with the PPVT-III (Dunn & Dunn, 1997). Results showed that monolinguals outperformed bilinguals in their L2 (English) at all ages and that the gap between monolinguals and bilinguals did not decrease with age. A further analysis was made on the type of receptive vocabulary (‘home words’ vs. ‘school words’) of a sub-group of the 6-year-olds (N = 161). Results showed that monolingual children scored significantly higher than bilinguals on English ‘home words’ but that the groups performed alike on ‘school words’ (Bialystok et al., 2010).

In an aforementioned study in Canada, Paradis (2011) found that home language use at home and the mother’s education level were not significant predictors for the children’s English receptive vocabulary, while age (in months), months of exposure to L2 and richness of the English environment were. Furthermore, Paradis found that the child’s use of English to family members was more closely associated with English vocabulary than the use of English to the child. The author speculated that this might be due to the parents’ low proficiency in English (Paradis, 2011).

In another abovementioned study from Canada, Thordardottir (2011) examined the relation between language exposure and receptive and expressive vocabulary development of 49 French-English-speaking bilinguals, 19 French-speaking and 16 English-speaking monolinguals (age range 52–69 months, all with high SES). All the bilingual children were exposed to both

---

31 These results need to be interpreted with care since there are unequal group sizes in the age groups. For example, in the 6-year-old group, there were 272 monolinguals and 458 bilinguals, while in the 10-year-old group there were 20 monolinguals and 15 bilinguals.

32 ‘School words’ included words related to school experiences, professions, animals or plants, shapes and musical instruments, while ‘home words’ included words that were unlikely to occur in a classroom context, such as words related to food and household as well as culture-specific items.

33 English richness environment outside of school was measured by asking the parents about their child’s amount of exposure to English through media, computer games, TV, books, organized activities and friends over the course of a week.
languages before the age of 3 but with varying proportions. The bilingual children were divided into three exposure groups: mostly English exposure ($N = 16$), mostly French exposure ($N = 19$), and roughly equivalent amounts of exposure to both languages ($N = 13$). Results showed that there was a strong relationship between the proportion of exposure to a language and vocabulary scores in that language. However, the effect of exposure was stronger for expressive vocabulary than for receptive vocabulary. Additionally, 5-year-old children with equal exposure (40%–60%) to both languages scored, as a group, similarly to the monolingual children in their receptive vocabulary in both of their languages, but for expressive vocabulary, more exposure was needed to reach the monolingual level. Children with unequal language exposure performed considerably stronger in the language they were exposed to the most (Thordardottir, 2011).

Rowe (2012) investigated the role of quantity and quality of parental language input on 50 English-speaking monolingual children in the US. Parent-child language interactions (video-taped free-play interactions at home) were measured at age 18, 30 and 42 months, while the children’s receptive vocabulary (PPVT-III, Dunn & Dunn, 1997) was measured at age 30, 42 and 54 months, i.e. one year after each video-recording. Keeping SES (parental education) constant, Rowe explored which aspect of parental input (quantity or quality) predicted the child’s receptive scores. Results showed that at 18 months, parental input quantity was a positive predictor of the child’s vocabulary one year later, but that at 30 and 42 months, parent input quality (the use of a diverse and sophisticated vocabulary, as well as decontextualized language in narrative and explanatory utterances) predicted their receptive vocabulary one year later (Rowe, 2012).

Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina and Trueswell (2013) investigated whether parental contextualized speech input predicted child vocabulary 3 years later. Adult native English-speaking participants ($N = 218$) were shown video clips from Rowe (2012)’s recordings of parent-child interactions (when the children were around 18 months) and were asked to guess some of the parents’ target words, which were muted when presented. The percentage of correct guesses (i.e. the measure of how easily the meaning of the word could be inferred from the situational context) was used as a referential transparency score. Results showed that the quality of the parent input (as heard by the adult raters) correlated with later child receptive vocabulary (even after controlling for effects of input quantity). In other words, the referential transparency score (i.e. parental input quality and not input quantity) predicted the child’s receptive vocabulary three years later (Cartmill et al., 2013).

Cattani, Abbott-Smith, Farag, Krott, Arreckx, Dennis and Floccia (2014) examined the amount of English language exposure needed for 28–32-months-old English-speaking bilinguals ($N = 35$, with diverse L1s) to perform on par with English-speaking monolingual peers ($N = 36$) on various English
vocabulary measures. Results showed that in order for bilingual children (mean age = 2;6) to be as proficient in English vocabulary (as measured by the CDI, Hamilton et al., 2000 and British Picture Vocabulary Scale (BPVS III, Dunn et al. 2009) as their monolingual peers, an English language exposure of 60% or more was needed. From a small sample of the bilingual children who had Arabic as their L1 (N = 13), the authors found that exposure to English was a strong negative predictor of Arabic language performance as measured by an Arabic version of the CDI (Cattani et al., 2014).

Hoff, Core, Place, Rumiche, Señor and Parra (2012) investigated the expressive vocabulary knowledge of 47 Spanish-English-speaking bilingual children and 56 English-speaking monolingual children in the US. The parents had a high SES and were asked to fill in vocabulary checklists (English: CDI, Fenson et al., 1993; Spanish: IDHC, Jackson-Maldonoda, Thal, Fenson, Marchman, Newton & Conboy, 2003) in all of the children’s languages at three points in time (at ages 1;10, 2;1 and 2;6) as well as answer a questionnaire querying the children’s language input. Bilingual children were divided into three exposure groups: Spanish-dominant (less than 30% English exposure), balanced bilingual (around 50%-60%), English-dominant (more than 70% English exposure). Age was a significant predictor of English scores for all three bilingual exposure groups. The size of the difference between the monolingual and bilingual children’s vocabulary skills in English depended on how much language input the bilingual child heard. When it comes to the bilingual children’s Spanish vocabulary scores, all exposure groups significantly increased with age. Results also showed that the Spanish-dominant exposure group scored significantly higher than the balanced and the English-dominant groups (who did not score differently from each other) in Spanish vocabulary. The authors concluded that bilingual children acquire lexical knowledge at the same rate as monolinguals, but that this word learning process, like their language exposure, is divided between the two languages (Hoff et al., 2012).

In a follow-up longitudinal study, Hoff, Rumiche, Burridge, Ribot and Welsh (2014) studied the expressive vocabulary development of a subgroup of the children in Hoff et al. (2012) (all having high SES): two groups of Spanish-English-speaking bilingual children (with one native Spanish-speaking parent N = 15, and with two native Spanish-speaking parents N = 11) and one English-speaking monolingual group (N = 31). Expressive vocabulary was measured at age 22, 25 and 30 months with English and Spanish vocabulary checklists (CDI, Fenson et al., 2007; IDHC, Jackson-Maldonado et al., 2003). At age 48 months, English and Spanish expressive vocabulary was measured

---

34 All the children were tested with the BPVS III (Dunn et al., 2009) for receptive vocabulary, Object naming subtask (adapted English SETK-2) for expressive vocabulary, Oxford Communicative Development Inventory (Oxford CDI, Hamilton et al., 2000) for parental report of the child’s expressive and productive vocabulary, and additional language specific CDI checklists for the children’s L1.
via a picture naming task, the Expressive One Word Picture Vocabulary Test (EOWPVT; Brownell, 2000). Results regarding total vocabulary growth (as measured by the parental vocabulary checklists) showed that children from bilingual homes with two Spanish-speaking parents were gaining over the monolinguals. When comparing the vocabulary results of the two bilingual groups, children who had one English-speaking parent and one Spanish-speaking parent had consistently larger English expressive vocabularies than children with two Spanish-speaking parents (but consistently smaller English vocabularies than children from monolingual homes). Concerning the bilingual children’s Spanish expressive vocabulary growth, children with two Spanish-speaking parents increased their Spanish vocabulary skills more than children with only one Spanish speaking parent. The gain over time did not come at the expense of their English vocabulary development. Similarly, the relative decline in Spanish expressive vocabulary for children with one Spanish-speaking parent was not associated with any relative gain in their English vocabulary. As for the children’s English expressive vocabulary at age 48 months, results showed that monolingual children scored significantly higher than children with two Spanish-speaking parents but not higher than bilingual children with one English-speaking and one Spanish-speaking parent (who scored within the monolingual norm). For the bilingual groups, having one English-speaking parent speak more English over time benefited the child’s expressive English vocabulary. On the other hand, having two Spanish-speaking parents who increasingly spoke more English over time did not benefit their children’s English scores but rather lowered the children’s Spanish scores (Hoff et al., 2014).

In an abovementioned study, Buac, Gross and Kaushanskaya (2014) examined whether there was a relation between language exposure and the receptive and expressive vocabulary knowledge (in both languages) of 58 Spanish-English bilingual children (mean age 5–7). Buac et al. found a significant relation between English (L2) language exposure and expressive and receptive English vocabulary. For Spanish (L1) vocabulary, Spanish language exposure had a significant effect on expressive vocabulary but not on receptive vocabulary.

Gagarina, Armon-Lotem, Altman, Burstein-Feldman, Klassert, Topaj, Golcher and Walters (2014) studied the effect of age, input quantity and input quality on 90 Russian-German-speaking children (47–86 months) and 79 Russian-Hebrew-speaking children’s (54–84 months) expressive vocabulary in both of their languages in comparison to monolingual norms. Information about the children’s language background was gathered via parental interviews and questionnaires. Expressive vocabulary was measured using a noun-verb naming task (Kauschke, 2007). Results showed that bilingual children scored significantly lower in their L2 than their respective monolingual norms. Furthermore, age was positively correlated with L1 expressive vocabulary in both bilingual groups, while length of exposure to their L2 had no negative
effect on their L1 vocabulary. The amount of L1 spoken at home (input quantity) had a strong positive effect on the children’s L1 expressive vocabulary (it was the strongest predictor of competence in L1) and it had no significant effect on their L2 (German/Hebrew) except to a minor extent on the expressive vocabulary of nouns \( p = .043 \) but not on verbs (Gagarina et al., 2014).

In a large-scale study, Gathercole and Thomas (2009) examined the receptive vocabulary of Welsh-English-speaking children (age 3–11) in relation to the children’s language input. Information about the children’s language input was provided by the children in the form of a questionnaire (and from the parents when complementary information was needed). The children were divided into three age groups 4 (3;0–5;9), 7 (6;0–8;5) and 9 (8;6–11;6). Receptive vocabulary was measured with the BPVS in English (Dunn, Dunn and Whetton, 1982) and its adaptation to Welsh (Spencer, 2000). Results from 611 children showed that the children’s receptive vocabulary in Welsh was significantly affected by the child’s age and the language spoken to at home (children who lived in homes where only Welsh was spoken scored significantly higher than children where both English and Welsh and only English was spoken at home, in all age groups). Results from 240 children showed that for both Welsh and English, age and language spoken at home significantly affected the children’s receptive vocabulary in both languages. For Welsh, the effect of home language use was clearer for the older children (i.e. 8;6–11;6) when a longer (more complex) version of the receptive vocabulary test was used. For English, the language use at home had an effect on the older children, however, the school language seemed to play a greater role: Children who were hearing both English and Welsh on a daily basis in school scored significantly higher in English receptive vocabulary than children who attended Welsh-only schools (Gathercole & Thomas, 2009).

Gathercole, Thomas, Roberts, Hughes and Hughes (2013) studied the receptive vocabulary knowledge of 331 Welsh-English-speaking bilingual children and 96 English-speaking monolingual children (age 2–15). Receptive vocabulary was measured in both English (BPVS, Dunn et al., 1997) and Welsh (Prawf Geirfa Cymraeg, Gathercole & Thomas, 2007). The children were classified into four ‘home language groups’ based on the family’s language use at home: monolingual English, bilinguals but with only English at home, bilinguals with both English and Welsh at home, bilinguals with only Welsh at home. Furthermore, the children were classified into four age groups (2–3, 4–5, 7–8, and 13–15). Results showed that for both Welsh and English, receptive vocabulary was significantly affected by age and language use at home in all age groups. However, as children got older, the ‘home language groups’ gradually became less distinguishable, meaning that early differences at young age in receptive vocabulary performance between the ‘home language groups’ were neutralized as the children got older (Gathercole et al., 2013).
Bohnacker, Lindgren and Öztekin (2016) investigated the expressive vocabulary (CLT, Haman et al., 2015) in the home language (L1) of two preschooler bilingual groups in Sweden (Turkish-Swedish, N = 40; German-Swedish, N = 38; aged 4;0–6;11) in relation to various background factors. The authors found no effect for age, SES (parental education) or MTI instruction on expressive vocabulary. They did, however, find a strong effect of language input from home, where children whose both parents spoke to each other and to the child in the home language scored significantly higher on vocabulary production than children whose parents did not do so. Additionally, children whose parents did not speak mainly the home language with them (or where only one parent did so) but were provided with additional home language input from other sources, for example friends, showed a clear positive difference in expressive vocabulary scores.

Balilah (2017) examined the vocabulary skills of 59 Arabic-English-speaking children (age 6–9) residing in Canada. The children were tested in receptive and expressive vocabulary in both Arabic and English (Arabic Receptive-Expressive Vocabulary Test, AREVT, El-Halees & Wiig, 1999; PPVT-4, Dunn & Dunn, 2007; Expressive Vocabulary Test, EVT, Williams, 2006). Results showed that age and richness of language environment outside of school predicted the children’s performance in Arabic (L1) receptive and expressive vocabulary. For English (L2), age of onset was the most predictive factor for expressive vocabulary whereas age (in months) and richness of English environment outside of school mostly predicted English receptive vocabulary performance.

Montanari, Akıncı and Abel (2019) studied the expressive vocabulary in both languages of 98 Turkish-German-speaking children (age 6;3–10;10) in Germany using the Wortschatz- und Wortfindungstest für 6- bis 10-Jährige, WWT 6–10, Glück, 2011). The authors found a steady shift towards L2 (German) dominance due to L2 being the language of school instruction. Furthermore, Montanari et al. found that a continuous use of the L1 (Turkish) by parents in the family had a positive effect on L1 expressive vocabulary but no negative impact on L2 vocabulary. The use of L1 with peers had a positive measurable effect on L1 vocabulary.

Öztekin (2019), whose study is part of the BiLI-TAS project, studied the receptive and expressive vocabulary of Turkish-Swedish-speaking children (N = 102, age 4;0–8;1) in Sweden using the CLT (Haman et al., 2015) in both of the children’s languages in relation to language exposure (based on parental estimates). Öztekin found that children who were exposed to 80% or more Swedish (L2) during the day scored significantly higher on Swedish production but also significantly lower on Turkish production. There was no significant difference concerning the receptive vocabulary scores. Furthermore, children whose both parents spoke mostly or only Turkish (L1) with them at home scored significantly higher on Turkish comprehension and production than children with other parental input patterns.
Sorenson Duncan and Paradis (2020a) examined the relationship between language input/output and maternal education (and the language in which that education took place in) with the syntax and expressive vocabulary of 89 English-speaking bilingual children (age 4;2–5;6, with diverse L1s) attending half-day English school programs in Canada. The children’s expressive lexical skills were measured on a story-telling generating task (ENNI; Schneider, Dubé, & Hayward, 2005) in English (L2) by calculating the number of different words (types) and total number of words (tokens). Results from the regression model showed that the child’s L2 exposure in school, maternal L2 (self-rated) fluency, mother’s language input as well as the child’s language output were significant predictors for the children’s L2 expressive vocabulary. In other words, the longer the children had been in (English) school, the more they heard English from their mothers (input), and the more the L2 was spoken by the child, the higher the number of word types produced in English by the child. Also, children who heard more L2 from their mothers used more different words in their narrations. The authors add, however, that the effect of maternal input on the child’s vocabulary might become more limited as the children start to attend full-day educational programs in their L2. Furthermore, Sorenson Duncan and Paradis found that mothers with higher levels of L1-based education were less likely to speak English (L2) with their child, while mothers with higher L2-based education were more likely to speak in English with their child. Hence, the mother’s level of education and the language in which that education occurred influenced relative maternal L2 fluency and L2 input to the child (Sorenson Duncan & Paradis, 2020a).

To conclude, studies seem to agree that there is a strong relationship between language exposure (input quantity) and lexical development of bilingual children where relative exposure predicts vocabulary. Although bilinguals learn at the same rate as monolinguals, their language learning and their language exposure are divided between the two languages (Hoff et al., 2012). The amount of language exposure needed to reach monolingual norms might differ for receptive and expressive vocabulary (Thordardottir, 2011). Language input in L1 seems to boost the child’s L1 usually without affecting the L2, which will be learned in school and in society (Montanari et al., 2019). Exposure to L2 seems to negatively impact L1 proficiency (Cattani et al., 2014). Some studies have also highlighted the importance of language richness (language quality, for example, input of diverse and sophisticated vocabulary, decontextualized and explanatory language) inside and outside the home to predict children’s vocabulary (Balilah, 2017; Paradis, 2011; Rowe, 2012).

Most studies only report vocabulary results for children at a group level. Studies that also report the performance of individual children (e.g. Thordardottir, 2011; Gathercole et al., 2013; Gagarina et al., 2014; Bohnacker et al., 2016; Öztekin, 2019) emphasize that some children performed differently than their age group despite having seemingly similar language exposure patterns.
In the present study, the effect of language input quantity (measured by parental estimates) is investigated in Arabic-speaking bilingual children. Several sources of language input will be investigated, including estimated daily language exposure and parent/sibling input to the child, but also language input through MTI attendance and parent-child shared book reading. The latter two measures of language input (in addition to SES) also tap into the qualitative aspect of input. However, no in-depth analyses of language richness have been done in the present dissertation. The present study will however present examples of children in Sweden in a cross-sectional study to illustrate the performance of individual cases.

3.1.4 Siblings

Another method to measure language exposure is through the language a child hears from the siblings (in contrast to language exposure from parents or language exposure at home). In what follows, selected studies that highlight the effect of language input from siblings (in terms of birth order) and its effect on vocabulary are summarized.

An American study by Bornstein, Leach and Haynes (2004) examined the vocabulary of monolingual English-speaking first-born children ($N = 55$) and that of their second-born sibling when each was aged 1;8 months. The children’s vocabulary knowledge was assessed using three different methods: maternal reports (interview/oral reports and vocabulary checklists (ELI; Bates, Benigni, Bretherton, Camaioni & Volterra, 1979; Bates, Bretherton & Snyder, 1988), observations from recordings of spontaneous speech (transcribed mother-child video dyads) and standardized oral test of the children’s expressive and receptive vocabulary (Reynell Developmental Language Scales, RDLS; Reynell & Gruber, 1990). Results showed that birth order was only significant in favour of the first-born sibling as per the maternal report, but not for the observations of the child’s spontaneous speech nor for the standardized vocabulary testing (Bornstein et al., 2004).

Quiroz, Snow and Zhao (2010) examined the expressive vocabulary of 51 Spanish-English-speaking children (age 4–5) in the US. Expressive vocabulary was tested in both languages using the Woodcock Johnson Language Proficiency Battery–Revised (WLPB–R) Picture Vocabulary Test in English (Woodcock, 1991) and Spanish (Woodcock & Muñoz- Sandoval, 1995). The authors found that child language use with older siblings was positively associated with the children’s productive vocabulary scores in that language and negatively associated with the child’s vocabulary scores in the other language (Quiroz et al., 2010).

In an aforementioned study, Paradis (2011) found that having an older sibling or not did not predict higher English L2 receptive vocabulary scores (PPVT-III, Dunn & Dunn, 1997) of 169 English-speaking bilinguals (mean age 5;10, with diverse L1s) in Canada. Recall, however, that the families were
newcomers to Canada and had relatively low exposure to English (3–62 months, mean = 20 months).

In a study of 31 Russian-Hebrew-speaking and 18 English-Hebrew-speaking bilingual children (age 4;4–6;1), Armon-Lotem, Joffè, Abutbul-Oz, Altman and Walters (2014) found that first-borns in the Russian-Hebrew-speaking group (and not in the English-Hebrew-speaking group) scored higher in their L2 Hebrew (Goralnik, 1995) than later-born children (Armon-Lotem et al., 2014). The results are hampered by the fact that the birth-order distribution in the groups differed.

Bridges and Hoff (2014) examined whether older school-aged siblings influenced the language environment and language development of toddlers in bilingual homes in the US. In two studies, parents filled in a questionnaire estimating the amount of time English was spoken at home and between the siblings. Information regarding the children’s English vocabulary knowledge was gathered via CDI checklists. In the first study, Bridges and Hoff found that in a group of 60 bilingual children (age 16–30 months, with diverse L1s), children with older siblings (N = 26) had significantly higher English (L2) vocabulary scores than those without older siblings. Furthermore, children who were reported to hear mostly English from their older siblings had higher English vocabulary scores than those who heard English in addition to another language from their older siblings. In a second study, 27 Spanish-English-speaking bilingual children were tested at age 22 months and again at 30 months. For language exposure, the study showed that having an older sibling increased the use of English (L2) in the home and that older siblings spoke more English to their siblings than what the mothers did, at both testing times. As for language development, toddlers who had older siblings were more advanced in English and less advanced in Spanish (L1) than children who did not have older siblings. Conversely, toddlers who did not have older siblings were more advanced in Spanish than they were in English. Hence, having an older sibling influenced the language development of younger children and affected the distribution of their vocabulary knowledge in their L1 and L2 (Bridges & Hoff, 2014).

In a larger study, Rojas, Iglesias, Bunta, Goldstein, Goldenberg and Reese (2016) investigated the expressive language skills of 224 Spanish-speaking English learners (age 4;10–6;9) in kindergarten in the US in relation to language use with different interlocutors (siblings and peers) and maternal education. Information about language use was gathered via parental question-

---

35 The children’s language proficiency was assessed only in their L2 (Hebrew) by testing the children’s vocabulary, listening comprehension, sentence repetition, sentence production, narratives and pronunciation.
naires and the children’s expressive language was assessed via narrative samples (*Frog, Where Are You?*; Mayer, 1969). Expressive language skills in Spanish and English were associated with whom the child interacted with, and for both languages, siblings’ language use was more influential than parents’ language use. The findings of Rojas et al. highlight the importance of older siblings in shifting children’s language use from the minority language (L1) to the majority language (L2). Surprisingly, the amount of sibling input was not predictive of any language measure but it was the child’s output to the older sibling that predicted an increase in English and a decrease in Spanish vocabulary (as measured by number of different words and words per minute) (Rojas et al., 2016).

In a study conducted in Sweden by Kheirkhah and Cekaite (2018), sibling language use in five Persian-Swedish-speaking families was analysed via interviews, observations and video-recordings of home interaction during playtime and mealt ime. The children were 3–17 years old. Although the parents were reported to mainly speak the heritage language (Persian) at home, the Sweden-born siblings mainly spoke Swedish. Siblings were thus viewed as the main source for majority language use in the family and, as a result, being the main contributors to family language shift towards Swedish (Kheirkhah & Cekaite, 2018).

In an aforementioned study, Montanari et al. (2019) found that children who spoke L2 (German) with their younger siblings tended to show L2 expressive vocabulary dominance more than children who spoke only L1 or those who spoke both L1 and L2 or children who did not have any younger siblings. No effect of L2 language use with older siblings was found for L2 expressive vocabulary.

Paradis, Soto-Corominas, Chen and Gottardo (2020) explored the receptive vocabulary knowledge of recently arrived Arabic-English-speaking children (*N* = 133, age 6–13) in Canada with a mean length of residency of 23 months (range 2–37 months). The children were tested for vocabulary comprehension in both of their languages (English: PPVT, Dunn & Dunn, 2007; Arabic: ALAB; Assadi, Shany, Ibrahim, Khateb, & Ben Simone, 2015) and background information was gathered via parental questionnaires administered via interviews. Results showed that more English use with siblings was associated with better English receptive vocabulary, but not worse Arabic receptive vocabulary. The fact that use of English between siblings was not negatively associated with Arabic vocabulary results was explained by the relative short time of residency in Canada and was predicted to change as length of residency increased (Paradis et al., 2020).

---

36 For each of the child’s two narratives, one in Spanish and one in English, the following expressive language measures were calculated: utterance length in words (mean length of utterance in words, MLU), lexical diversity and productivity (number of different words, NDW) and verbal fluency (words per minute, WPM).
In Lebanon, 112 Arabic-English/French-speaking toddlers (age 8–16 months) participated in a study by Daccache, Messarra, and dos Santos (2020). The parents were asked to fill in a Lebanese version of the CDI vocabulary checklist (IDC-L trilingue, Fenson et al., 1993) where they estimated their child’s receptive and expressive vocabulary knowledge in three languages (Arabic, English and French). Results showed that there was no significant difference in comprehension or production scores between first-born children ($N = 60$) and children who had an older sibling ($N = 52$) (Daccache et al., 2020).

An aforementioned study by Sorenson Duncan and Paradis (2020b) investigated the effect of older siblings’ language use on 113 English-speaking bilingual children (age 4;10–7;2, with diverse L1s) residing in Canada. All the children had at least one older sibling (mean number = 1.6, range 1–4, mean age of 12;3) and had a mean length of exposure to English at school of 16.7 months. Results showed that there was a very strong correlation between the children’s relative $L2$ output and the relative input the children received from their older siblings. Children who had more English input from their older siblings used more $L2$ English word types when narrating and scored higher on the receptive vocabulary test (Sorenson Duncan & Paradis, 2020b).

In sum, several studies agree that children tend to hear more $L2$ from their siblings than from their parents (e.g. Bridges & Hoff, 2014; Kheirkhah & Cekaite, 2018; Rojas et al., 2016; Sorenson Duncan and Paradis, 2020b). Hence, having a sibling contributes to shifting the family’s language dynamics towards speaking more of the majority language ($L2$) at home. On the other hand, short length of exposure to the majority language (associated with the family’s short length of residency) may contribute to the upkeep of the children’s $L1$ rather than shifting to $L2$ (Paradis, 2011; Paradis et al., 2020).

Studies seem to disagree on whether the child’s birth order significantly affects lexical knowledge in either language. Some studies find that children who have an older sibling have a more advanced $L2$ (e.g. Bridges & Hoff, 2014; Paradis et al., 2020), yet other studies find no difference between siblings’ birth order and their vocabulary knowledge (Daccache et al., 2020). However, the age of the children varied from one study to another, as did the vocabulary measure (see above). Bornstein and colleagues (2004) found that birth order was only a significant factor favouring the first-born child when maternal reports were used to assess the children’s vocabulary knowledge, but not when actual vocabulary testing was done. Finally, Sorenson Duncan and Paradis (2020b) emphasize the relative quantity of $L2$ input (varying degree of $L2$ exposure) from older siblings, rather than the mere existence of older siblings in a bilingual child’s life. In the present dissertation, the effect of language input from siblings is analysed in terms of estimated exposure (‘mostly Arabic’ and ‘mostly Swedish’) and in terms of birth order (first born vs. having older sibling(s)).
3.1.5 Parent-child joint book reading

Language input at home is more than mere conversational interaction. Parents may, for example, engage in literacy activities with their children in the form of parent-child joint, or shared, book reading. Shared book reading is believed to invite more parental language input, more verbal response output from the child and more verbal interaction between the two than other activities performed at home, like singing or playing with toys (Clemens & Kegel, 2021). However, studies of different bilingual (and monolingual) groups show contradictory results as to whether shared book reading measurably boosts children’s vocabularies in both language modalities, comprehension and production. Below an overview is given of selected studies on joint book reading and its effect on children’s vocabulary knowledge (in both monolingual and bilingual children).

Evans, Shaw, and Bell (2000) studied the influence of home literacy activities on early literacy skills in English in 66 children (age 5;5–6;8) in Canada. English was the children’s first language (for all except one child) and their preferred language. The children’s receptive vocabulary was measured with the PPVT-Revised (Dunn & Dunn, 1981), and information on home literacy activity was gathered via parent interview, child interview and home observations. After taking into consideration the child’s age and the parents’ education level, joint reading did not have any measurable effect on receptive vocabulary (Evans et al., 2000).

Frijters, Barron, and Brunello (2000) examined whether parent-child joint book reading had an effect on the receptive vocabulary (PPVT-R, Dunn & Dunn, 1981) of English-speaking children (N = 92, age 63–76 months) in Canada. Information regarding home literacy was gathered from parental questionnaires. Results showed that parent-child joint reading frequency (i.e. ‘How many times a week do you read to your child?’) explained 8% of the variance in the children’s oral receptive vocabulary, but that adding additional measures that assessed sources of home literacy experience (such as number of books read/possessed) doubled the variance accounted for (Frijters et al., 2000).

Weizman and Snow (2001) examined whether sophisticated37 home lexical input (including joint book reading) to 53 English-speaking children (age 5;3–5;8) affected the children’s receptive vocabulary at kindergarten age (5 years) and later in second grade (at around age 7). The children were not identified as monolinguals or bilinguals but were said to represent a range of socio-economic, cultural racial and educational backgrounds in the US. The children’s receptive vocabulary was measured with the PPVT-Revised (Dunn & Dunn, 1981). Results showed that, while there was a moderate association between mother-child joint book reading and the child’s receptive vocabulary, the

---

37 The authors define sophisticated lexical input as going beyond the 3000 most common English words and their inflected forms.
strongest correlations were found with conversations during *playtime* and *mealtim*e. Weizman and Snow concluded that lexically rich, naturally occurring conversations in early life affected the children’s receptive lexical knowledge at kindergarten age and up to second grade more than merely joint book reading.

In a longitudinal study by Sénéchal and LeFevre (2002), the effect of parent-child shared reading activity on receptive vocabulary was examined for 168 English-speaking children in Canada. The children were recruited at age 4–5 and were then retested in grade 1 and grade 3. Receptive vocabulary knowledge was measured with the PPVT-Revised (Dunn & Dunn, 1981) and home literacy experience (storybook reading/exposure) was measured via a questionnaire. Sénéchal and LeFevre found that home shared book reading during the preschool years correlated positively with the children’s receptive vocabulary, and that in time, these receptive language skills facilitate reading fluency in grade 3 (Sénéchal & LeFevre, 2002).

In a longitudinal study by Deckner, Adamson and Bakeman (2006), 55 English-speaking children in the US were tested between 18–42 months to determine the effects of home literacy on the children’s expressive and receptive vocabulary. The children’s expressive vocabulary (EVT, Williams, 1997) and receptive vocabulary (PPVT-III, Dunn & Dunn, 1997) were assessed at 30 months and 42 months. Results showed that parents’ literacy efforts during the child’s young age made “meaningful and lasting contributions” (p. 39) to both expressive and receptive vocabulary, even after controlling for initial vocabulary differences between the children (tested at age 18 months) (Deckner et al., 2006).

A study of 68 4-year-old Dutch-speaking monolingual children by Leseman, Scheele, Mayo, and Messer (2007) examined the effect of home literacy activities on the children’s receptive vocabulary skills. Vocabulary was measured via a picture identification test and information about home literacy activities was gathered via an interview with the mother. When dividing home literacy into oral and written forms, Leseman and colleagues found that *oral forms* of home literacy contributed in particular to the lexical level (including receptive vocabulary), whereas *written forms* of home literacy contributed in particular to the morpho-syntactic and textual level of developing academic language. For the children’s receptive vocabulary in particular, correlations showed that parent-child *talking at home* was more strongly associated (larger effect size) with receptive vocabulary than *joint book reading* (Leseman et al., 2007).

---

38 It was not explicitly stated whether the children were monolinguals or not, however, the authors mentioned that the children came from English-speaking homes and that “most [children] were White” (p.448). The children were from upper and middle class.

39 An example of *oral forms* of home literacy is ‘talking about personal experiences and topics of general interest’, while *written forms* of home literacy include ‘reading narratives and information books to the child’ (Leseman et al., 2007, p. 351).
In the US, 51 Spanish-English-speaking children (average age 5;2 years, SD = 4 months) participated in a study by Quiroz, Snow and Zhao (2010) which investigated whether interactions during book reading strengthened expressive vocabulary across both languages and not only in the language in which book reading occurred. Background information about home literacy activities was gathered via a phone interview, and the children’s expressive vocabulary was tested with the Woodcock Johnson Language Proficiency Battery–Revised in English (Woodcock, 1991) and Spanish (Woodcock & Muñoz-Sandoval, 1995). Results showed that for both languages, the mother’s reading was positively associated with the child’s expressive vocabulary in that particular language, but negatively associated with the other language’s vocabulary scores. In other words, children who were being read to in English scored better in English vocabulary production but worse in Spanish production. Interestingly, the mothers’ labelling questions (such as asking ‘What do you call this little animal?’) during joint book reading were positively associated with the children’s vocabulary scores in both Spanish and English (Quiroz et al., 2010).

In the Netherlands, 58 monolingual Dutch children, 46 bilingual Moroccan-Dutch children and 55 bilingual Turkish-Dutch children (all aged 35–43 months) participated in a study by Scheele, Leseman and Mayo (2010). The study investigated the relationship between home-language learning activities and receptive vocabulary. The parents filled in a questionnaire estimating their child’s experience with language(s) at home through reading activities and oral language interactions. The children’s receptive vocabulary skills were tested via picture identification tasks. Scheele and colleagues found that for the Dutch-speaking monolinguals, frequency of joint book reading (as well as storytelling and conversations, but not television watching and singing) correlated with the children’s receptive vocabulary. As for the bilingual groups, language input through joint book reading in neither L1 nor L2 correlated with the children’s receptive vocabulary knowledge in any of their languages while storytelling and conversations in L1 and L2 did. This result was explained by the fact that the monolingual Dutch-parents were reported to read much more to their children than what the Moroccan-Dutch and Turkish-Dutch parents did.

In her doctoral dissertation, Sim (2012) examined whether home reading had any effect on the English vocabulary of 80 children (age 4;9–6;3) in Australia. The children included both monolinguals and bilinguals, and they were tested for their expressive (noun) vocabulary (HPNT, Fisher & Glenister, 1992) and receptive vocabulary (PPVT, Dunn & Dunn, 1997). The children were divided into three groups: one control group and two joint-reading intervention groups. The parents in the two groups were required to read regularly

---

40 English was reported to be the only home language of 79% of the participants, while 21% were reported to speak an additional language at home.
(preferably 3 times/week) during a period of eight weeks. Results showed that children whose parents participated in either of the intervention joint-reading groups scored significantly higher on expressive vocabulary but not on receptive vocabulary than the control group (Sim, 2012).

In her doctoral dissertation, Wang (2014) investigated whether there was a relation between mother-child reading quality (instructional or affective) and receptive vocabulary. Mandarin-speaking monolingual children in Taiwan ($N = 103$, age $3;0–6;7$) were tested for their receptive vocabulary on the Chinese PPVT-Revised (Dunn & Dunn, 1981). Results showed that children whose mothers used more *instructional quality* during shared book reading had significantly higher receptive vocabulary scores than children whose mothers used *affective quality* book reading techniques (Wang, 2014).

Prevoo, Malda, Mesman, Emmen, Yeniad, Van Ijzendoorn and Linting (2014) studied the Dutch *expressive* vocabulary (Expressive One Word Picture Vocabulary Test, Brownell, 2000) and Turkish *receptive* vocabulary (Turkish translation of the PPVT, Dunn & Dunn 2007) of 111 Turkish-Dutch-speaking children (mean age $6;1$) in the Netherlands and found that the higher the SES, the more mothers read in Dutch, the societal language, and the more the children had Dutch books at home. Children whose parents read to them in either language (joint book reading), scored significantly higher in that respective language.

In a longitudinal study in the US, Demir-Lira, Applebaum, Goldin-Meadow, Levine (2018) tested 55 English-speaking children (at age $14, 18, 26$ and $30$ months) for whether their parents’ early joint book reading activity had any effect on the children’s later language and literacy outcomes. The children were reported to come from families where English was the primary home language without excluding potential bilingual children. Receptive vocabulary was tested with the PPVT (Dunn & Dunn, 2007) when the children attended second grade (at around age $7$). Results showed that early parent-child joint book reading predicted the children’s later receptive vocabulary knowledge. The authors also conclude that parents’ language use during book reading was of higher quality and quantity (vocabulary diversity and syntax

---

41 The parents also participated in one reading training session, focusing on either *dialogic reading* or *dialogic reading with addition of print referencing*. See Sim and Berthelsen (2014) for brief examples from both intervention session types.

42 *Instructional* reading quality included whether the mother asked book-related questions/requests (for example, posing yes/no questions or tag questions), gave book-related feedback (for example, praised the child’s engagement) and gave story structure comments (for example, talking about the story characters). *Affective* reading quality included the mother’s verbal reading expression (for example, imitation of character voices), contact with the child (for example, the child sitting on the mother’s lap), and the mother’s enjoyment being with the child.

43 During each of the 4 data collection visits, Demir-Lira et al. (2018) recorded naturally occurring interactions (90 minutes) between parent and child that included shared book reading activities and other, non-reading related, activities. The study controlled for the quantity of talk parents provide outside the book reading activity (during the recorded home visits), the families’ SES (parental education and income), and the children’s early language skills.
complexity) as measured via type-token ratio than the parents’ language use outside book reading contexts (Demir-Lira et al., 2018).

In a study in the US, Breitfeld, Potter and Lew-Williams (2021) measured the receptive vocabulary (PPVT, Dunn & Dunn, 1997) of 39 English-speaking monolingual children (age 4;6–5;6) after being read a picture book story to by their parents. Results showed that children were able to learn new vocabulary from picture book reading sessions irrespective of whether the parents’ reading intention was to teach new words or to bond with the child. Additionally, the study showed that children who already had large vocabularies learned new words more easily than children with smaller vocabularies (Breitfeld et al., 2021).

In sum, most studies on monolingual children (except Evans et al., 2000) seem to find a positive relation between parent-child joint book reading and receptive vocabulary irrespective of the parents’ reading intention, be it informative or for bonding (Breitfeld et al., 2021). Some studies found effects of reading type (instructional versus affective, Wang, 2014) on receptive vocabulary while others put more emphasis on additional literary measures (such as number of books read, Frijters et al., 2000). Comparing joint book reading to parent-child conversation at home, mixed results were found regarding which activity had a stronger association with children’s receptive vocabulary. Demir-Lira et al. (2018) found that joint reading activities contained more diverse vocabulary than parents’ language outside of book reading contexts while Leseman et al., (2007) and Weizman and Snow (2001) found a stronger benefit of parent-child conversation outside of book reading for the children’s receptive vocabulary. Joint book reading seems to additionally have an effect on children’s expressive vocabulary (Deckner et al., 2006).

As for bilingual children, far fewer studies on the effect of joint book reading exist. However, results showed that the frequency of joint book reading in either language was associated with gains in expressive vocabulary in that language (Quiroz et al., 2010). Prevoo et al. (2014) studied different vocabulary measures (L2 expressive vocabulary, L1 receptive vocabulary) but found a positive effect on both when joint book reading was performed in the respective language. However, Scheele et al., (2010) found that storytelling (in contrast to book reading, which was not very frequent in their bilingual groups) had a significant effect on the L1 and L2 receptive vocabulary of bilingual children in the respective language.

In general, the majority of research seems to agree that joint book reading has a significant positive effect on children’s receptive vocabulary; however, studies on the effect of joint book reading on bilingual children’s vocabulary are scarce. In the current dissertation, potential effects of shared book reading will be investigated in relation to the vocabulary development of Arabic-Swedish-speaking children (in both of their languages) and of the bilingual children in Lebanon (in Arabic).
3.1.6 Mother tongue instruction (MTI)

Inside the home of a bilingual family, there are plenty of opportunities for a child to hear and use the minority language. Outside the home, the situation might vary. Bilingual children who live in a society where the home language is a minority language might not be able to practise their L1 language outside of the home, or they may have this opportunity, depending on the number of L1-interlocutors or interaction possibilities. One opportunity for a bilingual child to hear her/his mother tongue outside the home, and in an educational setting, is through mother tongue instruction (MTI). In Sweden, MTI is usually provided by the municipality (or organized through private initiatives) and typically takes place once a week for about an hour.

In a study conducted within the BiLI-TAS project, Bohnacker, Lindgren and Öztekin (2016) examined whether attending MTI classes had any effect on the expressive vocabulary (CLT; Haman et al., 2015) of 40 Turkish-Swedish-speaking and 38 German-Swedish-speaking bilingual children (age 4:0–6:11). The parents answered a yes/no question to specify whether their child attended MTI at the time of testing.44 The children were divided into two groups accordingly. Neither for the Turkish nor for the German bilingual group was there any significant difference in vocabulary scores between children who attended MTI and those who did not (Bohnacker et al., 2016). Similarly, when more Turkish-Swedish-speaking children (N = 102, age 4;0–8;1) were included in the analysis in a larger study by Bohnacker, Haddad, Lindgren, Öberg, and Öztekin (2021), no measurable effect of MTI attendance was found on the children’s expressive and receptive Turkish vocabulary. Bohnacker and colleagues also investigated the effect of MTI on the Arabic-Swedish-speaking children in the present dissertation and found an effect of MTI on Arabic vocabulary (CLT) scores. They also established that Arabic children received double the number of hours of MTI than the Turkish group did. The effect of MTI will be further investigated in the present dissertation in relation to other language input factors.45

In another study conducted in Sweden that includes slightly older children, Ganuza and Hedman (2019) examined whether attending MTI had any effect on the vocabulary knowledge of 120 Somali-Swedish-speaking bilinguals (age 6–12). The children’s receptive vocabulary was tested with a Somali version of the PPVT (Dunn and Dunn, 1997) which was developed for the study, a measure of vocabulary depth (antonyms, hypernyms and synonyms in both languages) with a tool developed for the study) in addition to testing the children’s reading comprehension. Results showed that the group of children who

44 For Turkish, 15 out of 40 children attended MTI and for German, 22 out of 38 children did.
45 It should be noted, however, that the children in these studies (and in the present study) are relatively young and therefore may not have attended many hours of MTI attendance. Hence, the ‘real’ effect of MTI might not have been visible (as tested for lexical knowledge through the CLT) at this age.
attended MTI for more than one year \( (N = 96) \) outperformed the children who had no or limited MTI attendance \( (N = 24) \) on reading comprehension, with a weaker impact on the children’s vocabulary proficiency. Ganuza and Hedman explain that the children in the study had access to Somali outside of MTI and that their receptive vocabulary or lexical depth may be less affected by attending MTI than by other school-related competencies (e.g. literacy) (Ganuza & Hedman, 2019).

In Germany, Montanari et al. (2020) examined the effect of MTI attendance on L1 receptive and expressive vocabulary (using the WWT 6–10, Glück, 2011) of Turkish-German-speaking children and Russian-German-speaking children (all aged 6–10). Montanari et al. found different effects on the two bilingual groups. Russian-German-speaking children who attended MTI scored higher than Russian-German-speaking children who did not, whereas no such differences were found between the Turkish-German-speaking children who attended MTI and those who did not. The authors speculate that the reason why MTI effects were different between the two bilingual groups might be the ‘heterogeneity of the quality and quantity’ of the MTI classes (Montanari et al., 2020, p. 166).

As can be seen, there is not enough research on the effects of MTI attendance on the lexical skills of bilingual children in Sweden.\(^{46}\) Hence, in the current dissertation, MTI attendance will be investigated for its potential effect on the lexical knowledge of Arabic-Swedish-speaking children in relation to other background factors.

3.1.7 Socio-economic status (SES)

The relationship between socio-economic status (SES) and vocabulary is believed to be related to the quality and quantity of the language input that a child receives and that in turn affects the child’s vocabulary development. SES can be operationalized in different ways, for instance through education level (of the mother, father or both), occupation, family income, area of residency or a combination of these measures. Children’s vocabulary development may be affected by SES-related factors, where, for example, high-SES families may be able to provide their children with better, higher quality education opportunities, while low-SES families might live in unprivileged environments often associated with low safety levels with higher levels of stress and instability (Meir & Armon-Lotem, 2017). Other aspects may be related to parental input quality and include differences in quality of parental speech (Rowe, \(^{46}\) In a study of older children (age 14–15, \( N = 22 \)), Walldoff (2017) examined the heterogeneity of handwriting, spelling and production of basic MSA morphosyntax in Arabic-Swedish-speaking children, comparing those who attended vs. those who did not attend Arabic MTI class, for a written translation task from Swedish to Arabic (MSA).
2012) or amount of literacy activities and number of books at home (Prevoo et al., 2014).

SES in monolingual children has been shown to greatly affect vocabulary (Hart & Risley, 1995), where SES is believed to be directly related to parental language input (Scheele et al., 2010). For bilingual children, several studies have found an effect of SES on vocabulary in the majority language but not on the minority language (e.g. Buac et al., 2014; Leseman, 2000) while others have found some effect on the minority language where children with low SES scored better than children with high SES on some vocabulary measures (Cobo-Lewis et al., 2002b). Below are short reviews of selected studies that have studied the effect of SES on children’s lexical knowledge.

In a famous longitudinal study in the US, Hart and Risley (1995) examined the vocabulary development of 42 English-speaking monolingual children in relation to the amount of language input they received at home. The authors visited the families for 2.5 years from when the children were approximately 12 months old and recorded speech samples for 1 hour every month. Hart and Risley found that SES (parental occupation and income) and amount of parental speech addressed towards the child and the child’s expressive vocabulary correlated. High-SES children were receiving much more language input than children of low SES. Although SES showed a relation with the child’s later vocabulary, particular language qualities of the parents’ speech addressed to the child (language input quality) affected the outcomes more. Hart and Risley documented that parental verbal engagement did not only vary between the low and high-SES groups, but also within the SES groups where different families provided different kinds of input, for example, concerning the number of words and use of questions, commands and prohibitions, irrespective of SES. Children who were not exposed to rich verbal engagement at a young age were more likely to lag behind their peers (in preschool and school age) who were exposed to richer verbal engagement and as a result had a steeper vocabulary growth curve (Hart & Risley, 1995).

Leseman (2000) examined the longitudinal development of expressive and receptive vocabulary of 31 Turkish-Dutch-speaking children of low-SES, 31 Dutch-speaking monolingual children of low-SES families (working class) and 46 Dutch-speaking monolingual children of high-SES families at age 3;2, 3;8 and 4;2. The children’s vocabulary was measured via a Turkish and Dutch receptive and productive vocabulary test (Narrain & Verhoeven, 1993). Results showed that for the bilingual children, SES (parental education) had a significant effect on the children’s expressive and receptive vocabulary in Dutch but not in Turkish at all testing occasions. The bilingual children showed slower development and performed significantly lower in Dutch vocabulary than the low-SES monolingual children, even at the final testing. The

---

47 The bilingual children’s Dutch expressive vocabulary could not be tested at the first testing due to limited knowledge.
bilinguals’ score in Turkish was similar to the low-SES monolinguals’ score in Dutch.

In a large-scale aforementioned study by Cobo-Lewis, Pearson, Eilers and Umbel (2002a, 2002b), 952 elementary-school students (high-SES, \( N = 485 \) and low-SES, \( N = 467 \)) were tested with four different oral lexical tests in both of their languages. The participants included both Spanish-English-speaking bilinguals (\( N = 704 \)) and English-speaking monolinguals (\( N = 248 \)). Socio-economic status (parental occupation) was significantly related to the children’s vocabulary scores in the English oral lexical tasks, where high-SES children scored significantly higher than low-SES children in the majority language (English). Families with high-SES also spoke more English at home than those with low-SES. As for the effect of SES on Spanish, the minority language, there was a significant effect on only two out of the four oral lexical tasks where children with low-SES children scored higher than high-SES peers in Spanish. Hence, the effect of SES on the minority language was not extensive on all vocabulary measures (Cobo-Lewis et al., 2002a, 2002b).

Hoff (2003) examined the effect of SES and maternal speech on the expressive vocabulary of 2-year-old English-speaking monolingual children in the US. The families were divided into high-SES (\( N = 33 \)) and mid-SES (\( N = 30 \)) based on the mothers’ education (university education vs. high school). Video recordings from 2 different occasions (10 weeks apart) were transcribed and used to measure the mother’s speech and the child’s expressive vocabulary (the number of word types in a 90-utterance speech sample). Results showed that high-SES children’s productive vocabulary had grown more than that of mid-SES children. This was explained by the properties of the maternal speech addressed to the child. Although high-SES mothers produced more utterances, more word types and word tokens, and had higher MLUs than mid-SES mothers, only the mother’s MLU significantly affected the child’s expressive vocabulary in the final regression analysis. In other words, children who heard longer utterances from their mothers built expressive vocabulary at a faster rate than children who heard shorter utterances. Hence, although mothers’ talk to their children differs in relation to SES, the particular language property of MLU contributes to the child’s expressive vocabulary. Mothers who speak in longer utterances also use a richer vocabulary, giving their children more opportunities to hear more different words and consequently learn more different words. (See Hoff, Laursen and Tardif (2002) for an extensive overview on the relation between SES on parental speech.) Hoff suggests that we should not focus on the effect of SES per se and instead put emphasis on the parents’ verbal engagement with their child serves as a predictor to the child’s vocabulary development (Hoff, 2003).

In an abovementioned study in the Netherlands, Scheele, Leseman and Mayo (2010) found that SES (parental education and current employment) was related to the qualitative language input by the parents and had a strong effect on the receptive vocabulary of Dutch-speaking monolingual children.
However, for the Moroccan-Dutch children, SES was not related to L1 (Tari-fit-Berber) receptive vocabulary but to L2 (Dutch) receptive vocabulary, while for the Turkish-Dutch-speaking group, SES was not related to receptive vocabulary performance, neither in L1 (Turkish) nor L2. The authors noticed that the Moroccan-Dutch families provided more L2 input at home in comparison to the Turkish-speaking families who had greater access to oral and literacy materials in their L1, something which the Moroccan families lacked. The authors cite Backus (2005) who emphasizes the importance of language maintenance amongst the Turkish immigrant community. Hence, this difference in attitude towards language maintenance between the Turkish-Dutch and Moroccan-Dutch families might have played a role in the L1 input at home.

Armon-Lotem, Walters and Gagarina (2011) also examined the impact of SES (parental education and occupation) on the expressive vocabulary (Kauschke, 2007) of 78 Russian-Hebrew- and 65 Russian-German-speaking children (age 3;11–7;2). The authors found no effect of SES for Russian-Hebrew L1 and L2 vocabulary, but a strong correlation of SES with vocabulary in both languages in the Russian-German-speaking children. Unfortunately, the two cohorts differed in SES where the Russian-German-speaking parents presented a wide range of SES, while the Russian-Hebrew-speaking group was more homogeneous in terms of the parents’ occupation and education.

In an aforementioned study, Rowe (2012) examined the receptive vocabulary (PPVT-III, Dunn & Dunn, 1997) of 50 English-speaking monolingual children (at age 30, 42 and 54 months) after having observed (video taped) free-play interactions one year earlier. The primary caregiver’s education (in years) was used as a proxy for SES. Results showed that SES significantly correlated with the children’s receptive vocabulary on all three testing occasions. Furthermore, there was a relation between SES and input quality and quantity, namely the higher the education level of the caregiver, the more words were spoken to the child and the more sophisticated the words were.

Calvo and Bialystok (2014) conducted a study of 175 6–7-year-old children (bilinguals of various L1s and English-speaking monolinguals) in Toronto, Canada, to examine whether SES (maternal education) and bilingualism affected the children’s receptive vocabulary (as measured with the PPVT-III, Dunn & Dunn, 1997) in the societal language, English.48 Results showed that children from high-SES (middle class) families scored significantly higher on receptive vocabulary than children coming from low-SES (working class) families, irrespective of whether the children were monolinguals or bilinguals. (The impact of SES was equivalent for both monolingual and bilingual children.) Furthermore, at a group level, monolinguals scored significantly higher than bilinguals irrespective of whether the children had a high-SES or low-

---

48 The children were divided into 4 groups: working class monolinguals (N = 20), working class bilinguals (N = 44), middle class monolinguals (N = 46) and middle class bilinguals (N = 65).
Calvo and Bialystok found no interaction with bilingualism and SES, meaning that the effect of each factor was independent of the other.

In an aforementioned study, Buac, Gross and Kaushanskaya (2014) examined whether the impact of SES (primary caregiver’s total years of education), exposure and the vocabulary knowledge of the primary caregiver had any effect on the vocabulary skills in both languages of 58 Spanish-English bilingual children (mean age 6.22). Results showed that SES significantly affected the children’s expressive and receptive vocabulary in the majority language (English) but not in the home language (Spanish).

A sample of 111 Turkish-Dutch-speaking children (age 5–6-year-olds, mean age 6;1) in the Netherlands participated in a study by Prevoo, Malda, Mesman, Emmen, Yeniad, Van IJzendoorn and Linting (2014) on the effect of SES (annual gross income and highest education level) on the children’s vocabulary. The children’s expressive vocabulary in Dutch (Expressive One Word Picture Vocabulary Test, Brownell, 2000) and their receptive vocabulary in Turkish (Turkish translation of the PPVT, Dunn & Dunn 2007) were measured. Data was gathered via a two-hour home visit, an interview with the mother, testing the child’s vocabulary and video observations. Results showed that high-SES status correlated with Dutch (L2) expressive vocabulary but not with Turkish (L1) receptive vocabulary. (Dutch receptive vocabulary and Turkish expressive vocabulary were not tested.) Furthermore, there was a negative correlation with SES and L1 language use (as measured on a 5-point scale), meaning that the higher the SES the less L1 was spoken at home by the mothers.

Meir and Armon-Lotem (2017) investigated whether SES (maternal education) impacted the expressive vocabulary of 88 Russian-Hebrew-speaking children (low-SES N = 44, mid-high-SES N = 44) and 32 Hebrew-speaking monolingual children (low-SES N = 16, mid-high-SES N = 16) aged 5;7–6;7. The children were only tested in the societal language, Hebrew, using the Goralnik Screening Test for Hebrew (Goralnik, 1995). Meir and Armon-Lotem found that SES significantly affected the vocabulary scores in the majority language for both bilinguals and monolinguals.

In an abovementioned study by Gagarina et al. (2017), SES (parental education and occupation) did not affect the German (L2) receptive and expressive vocabulary of 55 Turkish-German-speaking and 39 Russian-German-speaking children.

In aforementioned studies in Sweden, Bohnacker, Lindgren and Öztekin (2016) examined the effect of SES (parental education) on Turkish (L1) ex-

---

49 However, because of the large overlap between the scores of the bilinguals with high-SES and monolinguals with low-SES (large standard deviation of the bilingual high-SES group), on the individual level, bilinguals from high-SES may score as high as monolinguals of low-SES.

50 The primary caregivers (mothers N = 46, fathers N = 10, grandmothers N = 2) were reported to spend most time with the child on a daily basis.
pressive vocabulary (CLT, Haman et al., 2015) of 40 Turkish-Swedish-speaking bilingual children (age 4:0–6;11). The authors found no effect of SES on the children’s L1 vocabulary scores. Öztekin (2019) included a larger sample of Turkish-Swedish-speaking children in Sweden (N = 102, age 4:0–8;1) and also found no effect of SES on expressive or receptive vocabulary in Turkish and Swedish. Öberg (2020), who operationalized SES (parental education) as a continuous variable, also found no correlation between SES and the expressive and receptive vocabulary (CLT) in both Arabic and Swedish of 99 Arabic-Swedish-speaking children.

To sum up, many studies have found a significant effect of SES on bilingual children’s majority language (e.g. Buac et al., 2014; Calvo and Bialystok, 2014; Cobo-Lewis et al. 2002a; Leseman, 2000; Prevo et al., 2014), but no conclusive results regarding SES on the minority language. There could be several reasons for these inconsistent findings. One reason could be related to how SES was operationalized since different measures were used as a proxy for SES (for example, maternal or parental education level, occupation, or income). Another reason could be related to the different types of vocabulary tests that were used as well as what kinds of vocabulary measures were tested, whether expressive or receptive vocabulary (or both) were analysed. A third reason could be related to the difference in opportunities provided for low versus high-SES families in their respective country. In some countries, families with high-SES might have access to better learning experiences that stimulate vocabulary growth that low-SES families do not have access to. In other countries, differences in family SES might not be contributing so much to differences in language learning experiences, for example differences in (pre)school education opportunities (e.g. in Sweden as observed with the Turkish-Swedish-speaking children in Bohnacker et al., 2016 and Bohnacker et al., 2021).51 In the current study, the effect of SES (operationalized as parental education) on vocabulary development will be investigated for the bilingual children in the cross-sectional studies in Sweden and in Lebanon.

As an overall conclusion, earlier vocabulary studies point to general trends for some but not all child-internal and environmental factors. In the present dissertation, the above mentioned background factors that have been proven to influence children’s lexical development are taken into account. In the Swedish cross-sectional study, their effect is examined on both of the children’s languages (Arabic and Swedish), while in the Lebanese cross-sectional study, the effect of the background factors is examined for Arabic alone.

---

51 Bohnacker et al. (2021) also examined the effect of SES on the participants of the present study. The results are presented in Section 6.3.3.2 and are further compared to the effect of SES on the Arabic-bilingual children in Lebanon (Section 9.3.3.2).
3.2 Narrative macrostructure

There are numerous ways to define what a narrative, or story, is (see Stein and Policastro (1984) or Trabasso and Rodkin (1994) for an overview on different definitions of a narrative and what a narrative ‘should’ consist of). A concise summary is presented below.

A narrative is, according to Labov and Waletzky (1967), the presence of temporally connected sequences that consist of an onset, an unfolding and a resolution. Stein and Glenn (1979) emphasise that a narrative should have a setting (time and space) and a set of episodes, each structured around goal-directed action and an end which contains the outcomes of whether the goal was attained or not, as well as the reaction to the outcome. Importantly, these categories of information are to be connected by causal, temporal, or logical relations (Stein & Glenn, 1979). Stein and Policastro (1984, pp. 148–151) add that a story must include at least one animate protagonist and at least one form of causal sequence that includes a goal-directed action, with the optional omission of a reaction (outcome) and internal responses of the characters. However, the narrator of a story should keep in mind the listener’s perspective by clearly introducing and conveying information about the protagonists, their goals and their internal states (i.e. emotions) (see also Burris & Brown, 2014; Hudson & Shapiro, 1991; Trabasso & Rodkin, 1994). Hence, a narrative needs additionally to assume the point of view of the protagonist (Trabasso & Nickels, 1992). This implies that the narrator needs to have theory of mind competence, or the ‘awareness of internal states (e.g. attention, perception, desire, intention, emotion, knowledge, belief) and the ability to use this awareness in interpreting, explaining, and predicting the behaviour of self and others’ (Astington & Pelletier, 2005, p. 313).

In sum, a narrative should define its setting and include temporally, logically and causally connected events and actions of one or more protagonists whose goals and internal states are mentioned by the narrator in a manner that is understandable for the listener.

The ability to produce coherent and cohesive narratives is important for basic communication skills. Studying narrative competence is usually done at the level of macrostructure and/or the level of microstructure (e.g. Hickmann, 2004). These levels are often believed to be distinct yet interrelated within narrative competence (Justice et al., 2006). The analysis of narrative microstructure focuses on language-specific aspects (linguistic structures) such as measuring the production of conjunctions, noun phrases, dependent clauses (e.g. Justice et al., 2006) or the use of referential, temporal and causal linking devices (e.g. Berman, 1988). Narrative macrostructural analysis examines the overall hierarchical organization of narrative content with regard to story

52 The terms ‘narrative’ and ‘story’ are used interchangeably in the present dissertation.
structure (story grammar components) and episodic complexity (story complexity) (Westby, 2012). Macrostructure knowledge is often assumed to be largely language independent (e.g. Berman & Slobin, 1994; Pearson, 2002; Bohnacker, 2016; Paradis et al., 2011; Iluz-Cohen & Walters, 2012).

Macrostructural aspects of narrative performance are in focus in the present dissertation. Studying children’s narrative macrostructure is important since narratives are considered to be typical examples of decontextualized language that predict academic success and provide a basis for literacy development (e.g. Miller, 1990; Paul & Smith, 1993).

Studies have often found that narrative macrostructure strongly develops with age for pre-school and early school-aged children. For monolingual children, the production of story structure components and the combination of components to produce complex episodes develops between ages 3 and 9 with a sharp increase at around age 5 where children start to integrate information more fully into episodes (e.g. Berman & Slobin, 1994, pp. 46–84; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Mäkinen, 2014). Macrostructural skills depend more on age rather than on which language the story is told in, as observed by Berman and Slobin (1994) for monolingual English-, German-Hebrew-, Spanish- and Turkish-speaking children and adults telling the Frog story (Mayer, 1969). Differences were found between the age groups but not between the languages. Studies on several languages have found that the components attempts and outcomes are more commonly produced already at a young age whereas the production of goals and internal states appears later (e.g. Stein & Glenn, 1979; Trabasso et al., 1992; Trabasso & Rodkin, 1994; Soodla & Kikas, 2010). Similar findings have been observed for bilingual children where narrative production proved to be more dependent on age than on a specific language of bilingual children or language input factors (e.g. Akincı et al., 2001; Fiestas & Peña, 2004; Gutiérrez-Clellen, Simon-Cereijido & Wagner, 2008; Pearson, 2002; Uccelli & Páez, 2007). Hence, narrative macrostructure is often believed to be more reliant on general cognitive abilities (and less on microstructure) and similar (even universal) across languages of a bilingual child (the universality of macrostructure hypothesis). As a consequence, macrostructure is also often assumed to be similar in the two languages of a bilingual child. Furthermore, macrostructural skills of bilingual children may carry over between their languages (e.g. Pearson, 2002; Paradis et al., 2011).

Examining both narrative comprehension and production is important since children tend not to mention all story components when telling a story (e.g. Trabasso & Rodkin, 1994, p.103). In fact, several studies point out that young children have the ability to answer narrative comprehension questions (for example, questions probing goals) before even being able to tell a narrative (e.g. Stein and Glenn, 1979; Trabasso et al., 1992). Narrative comprehension is considered a necessity for narrative production, since in order to tell a story, a child must have an understanding of the story elements and of the relation they
have with each other (e.g. Astington & Pelletier, 2005, p. 327; Burris & Brown, 2014; Stein & Glenn, 1979; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Westby, 2012). Finally, narrative comprehension skill is often believed to be a prerequisite for school achievements since it is linked to literacy development and reading comprehension (e.g. Dickinson & Tabors, 2001; Paris & Paris, 2003; Lynch, van den Broek, Kremer, Kendeou, White & Lorch, 2008).

For the examination of narrative macrostructure production, numerous models have been developed (e.g. Labov & Waletzky, 1967; Stein & Glenn, 1979; Peterson & McCabe, 1983; Berman & Slobin, 1994; Westby, 1984). Some of these focus on quantifying elements of story structure and others on story macrostructural complexity. Studying story structure implies counting the number of story components produced. Story complexity, on the other hand, measures the combination of goals, attempts and outcomes within one episode where the combination of all three components is considered as the highest level of story complexity and an indicator of the ability to produce a coherent story that allows the listener to understand the motive and behaviour of the characters as well as the result of the action (Stein & Policastro, 1984; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992; Trabasso & Nickels, 1992; Westby, 2012, p. 211).

In the current study in Sweden, children’s narrative comprehension and production is analysed in Arabic and Swedish. Age development and language differences are investigated in addition to children’s narrative comprehension and production of story components.

The Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2012, 2019) is used in the current study as a tool to evaluate the narrative abilities of Arabic-Swedish-speaking bilingual children in both of their languages. MAIN contains four fictional picture sequence stories (Cat, Dog, Baby Birds, Baby Goats) that include different protagonists and events. Despite this, the stories are built to have a parallel overall story structure and episodic complexity organized around the protagonists’ goals. MAIN was developed using the goal-based story grammar model by Stein and Glenn (1979) and it contains a standardized protocol for how to administer and

---

53 Studies that have used other narrative elicitation tools may have assessed narrative macrostructure with other story grammar models than that used in MAIN. For example, studies who use the Frog stories (Mayer, 1969) often use the story grammar model proposed by Stein & Glenn (1979), Westby (1984), Trabasso, van den Broek & Suh (1989), Berman and Slobin (1994) or other (non-story grammar) models. Due to space limitations, such studies will not be mentioned here.

54 The story grammar model used in MAIN differs from the Stein & Glenn (1979) story grammar model in a few aspects. The Stein & Glenn story grammar includes a ‘major setting’ component (introduction of the protagonist(s)), whereas MAIN does not count character introduction as a scorable macrostructure component. Stein & Glenn’s model does not have a separate ‘goal’ component but the goal is instead part of an internal response category. MAIN, on the other hand, has a separate clear component for goals. The Stein & Glenn model has a initiating
The MAIN stories can be elicited in three modes: telling mode (story generation), retelling mode (hear the story then repeat it), model story (hear a different story then tell own story). MAIN is described in detail in Sections 4.2.2 and 4.5.2 in the Methods chapter.

Research using MAIN has flourished in the past years despite it being a relatively new narrative elicitation tool. The present chapter will summarize studies that have used MAIN with bilingual (and monolingual) children to examine their competence in narrative comprehension (Section 3.2.1) and narrative production (Section 3.2.2).

3.2.1 Narrative comprehension with MAIN

Narrative comprehension is the ability to identify the causal, hierarchical and thematic relations between several events and the ability to understand the overall plot or idea behind the story (Bohnacker & Gagarina, 2020, p. 4). Narrative comprehension can be investigated in different ways. As the present study used MAIN, the focus here will be on how narrative comprehension is tested with this instrument. Throughout the narration process, the MAIN picture sequence faces the child and not the experimenter. After having told the story, the pictures are put on the table so both child and experimenter are able to see the pictures when examining the children’s narrative comprehension. The comprehension questions are asked with shared common visuals in order not to tax memory unduly. The child is not asked to recall facts about the story from memory. Instead, the child is asked questions that require inferencing probing the child’s ability to understand the protagonists’ goals, internal state (as an initiating event or reaction) as well as theory of mind, all of which need to be inferred rather than ‘read off’ the visual picture sequence (Bohnacker & Gagarina, 2020).

Summarized below are studies that have used MAIN to study (bilingual) children’s narrative comprehension of macrostructure. Results on age development, task, elicitation mode and language effects are mentioned in addition to the children’s performance on specific comprehension questions.

In Slovakia, Kapalková, Polišenská, Marková and Fenton (2016) explored the narrative comprehension of 40 English-Slovak-speaking bilingual children (age 5–6) who attended English-speaking (pre)schools (since mean age of 38 months, L1 Slovak, L2 English). Two narrative modes were used: telling...

---

event component (an event that causes the main character to respond) which MAIN does not have. Both models share the components of setting (time, place), internal state as initiating event, attempt, outcome (or consequence as it is called by Stein & Glenn) and internal state as reaction. For an overview on other classical story grammar models, see Trabasso & Rodkin (1994).
mode (Baby Birds, Baby Goats) and retelling mode (Cat, Dog).\textsuperscript{55} Results showed that children scored equally well in both their languages and similarly across the languages on the different questions. Kapalková and colleagues found that questions targeting goals had the lowest accuracy rates in comparison to the other questions targeting internal states.

Maviş, Tuncer and Gagarina (2016) carried out two studies in the home language (L1 Turkish) of Turkish-German-speaking bilingual children in Germany. In the first study, 36 children (age 2;11–7;11) were asked to listen to a model story (Cat, Dog) and answer comprehension questions on it, then the children were asked to tell another story (Baby Birds, Baby Goats) and answer the story’s comprehension questions. The authors found age effects in both narrative elicitation modes.\textsuperscript{56} In the second study, 13 children (5;5–7;11) were asked to tell a story (Baby Birds, Baby Goats) and answer comprehension questions and were later asked to listen to a model story (Cat, Dog) and then retell the same story followed by comprehension questions. Results showed age effects where the older children (7;1–7;11) scored higher than the younger children (5;5–7;0).

In Italy, Roch, Florit and Levorato (2016) explored the narrative comprehension skills of 30 5–6-year-old and 32 6–7-year-old Italian-English-speaking sequential bilingual children (L1 Italian, L2 English). The children started to be exposed to English at around age 3 when they entered an English-medium preschool. Both telling (Baby Birds, Baby Goats) and retelling (Cat, Dog) modes of MAIN were elicited. Results showed that there were age effects and language effects for both modes. Younger children (who had only had around 2 years of exposure to English) performed better in their L1 than in their L2. No language differences were found for the older children. The authors conclude that age and length of exposure play a role in narrative comprehension competence.\textsuperscript{57}

In Sweden, Bohnacker (2016) examined the narrative comprehension of 52 English-Swedish-speaking bilingual children (5–7-year-old) in both languages using Baby Birds and Baby Goats in the telling mode. Bohnacker found no difference between the children’s two languages. Clear age effects were found where the 5-year-olds ($N = 19$) performed poorer than 6-to 7-year-olds ($N = 33$). Results showed that younger children were able to understand goals at a young age. Bohnacker concluded that children’s awareness of others’ states of mind and their ability to draw inferences are still developing at

\textsuperscript{55} Note that Kapalková et al. (2016) combined the narrative comprehension results for the telling and retelling modes which makes it difficult to compare the results with the present study. Furthermore, the authors did not include the last question in MAIN in their analyses.

\textsuperscript{56} Maviş et al. (2016) also found that the children scored significantly higher on the comprehension questions after the model story than after telling where they first had generated the stories themselves.

\textsuperscript{57} Similarly to Maviş et al. (2016), the children answered the comprehension questions after the retelling more accurately than after telling.
age 7 in contrast to the understanding of goals which was already well developed at age 5. Bohnacker also discussed the difference in children’s answers questions probing internal states, where younger children exclusively gave short-range explanations (i.e. relating the answer to local information) whereas some of the older children gave long-range explanations (i.e. relating the answer to the entire plotline rather than on one picture).

Lindgren (2018) found an age effect in the narrative comprehension skills for three Swedish-speaking groups of children, 72 Swedish-speaking monolinguals, 46 German-Swedish-speaking bilinguals, 48 Turkish-Swedish-speaking bilinguals all aged 4–6 living in Sweden. All four MAIN stories were elicited in the telling mode. For the bilingual groups, age development for L2 (Swedish) was more prominent than for L1 (German or Turkish), however no differences between the children’s two languages were found. For all three languages, performance was higher in Cat/Dog than in Baby Birds/Baby Goats. As for the individual comprehension questions, Lindgren found that the overall patterns for the accuracy rates was more or less similar between the three languages. In Cat/Dog, accuracy rates were quite high, except for one overall plotline question. For Baby Birds/Baby Goats, accuracy rates were lower. The two questions that were answered correct with the least frequency probed the children’s understanding of the protagonists’ theory of mind (emotion and rationale). Finally, Lindgren found a story effect where the monolinguals and German-Swedish-speaking bilinguals (but not the Turkish-Swedish-speaking bilinguals) scored higher on the comprehension of Baby Goats than Baby Birds.

Previous studies that have used Cat/Dog in the retelling mode and Baby Birds/Baby Goats in the telling mode (as initially intended by Gagarina et al., 2012, 2015, p.256), have found that children’s narrative comprehension scores in the retelling mode were higher than the comprehension scores in the telling mode (Maviş et al. 2016; Roch et al. 2016). In a paper by Bohnacker and Lindgren (2021), the authors compare the narrative comprehension skills of 52 English-Swedish-speaking bilinguals with 72 Swedish-speaking monolinguals from Bohnacker (2016) and Lindgren (2018), respectively. Bohnacker and Lindgren argue that comprehension of Cat/Dog is easier than Baby Birds/Baby Goats even when the stories are administered in the same mode (telling) since children as young as 4 already managed to score high on Cat/Dog. Likewise, at age 6, there were significant differences between the story pairs Cat/Dog and Baby Birds/Baby Goats. When analysing the response pattern of the individual questions of Baby Birds/Baby Goats, Bohnacker and

58 In a longitudinal study on 17 Swedish-speaking monolingual children, Lindgren (2019) found that age development for narrative comprehension from mean age 4;4 to 5;10 was significant but not from mean age 5;10 to 7;4. Lindgren found that children at 5;10 already scored around 80% correct on Baby Birds and Baby Goats (Cat and Dog were not analysed).
Lindgren found similar response accuracies for the monolinguals and bilinguals (as well as similar response accuracies for the two languages of the bilinguals). The authors also found that the questions with the highest response rate in both Cat/Dog and Baby Birds/Baby Goats probed the children’s understanding of goals which were close to ceiling by age 6. As a result, Bohnacker and Lindgren suggest that the difference in the performance between the story pairs is not related to goals but is related to questions probing internal state. Internal state questions, although formed with identical wording across the four stories, necessitate different inferences from one or several pictures, depending on the story. Bohnacker and Lindgren also discuss the children’s performance on questions probing internal states in the different story pairs where inferences from Cat/Dog can be made from one picture (short-range interpretations) whereas inferences from Baby Birds/Baby Goats are made from the whole story (long-range interpretations).

Lindgren and Bohnacker (2020) investigated the effect of age and expressive vocabulary (CLT) on narrative comprehension of the 46 German-Swedish-speaking children in Lindgren (2018). Results showed that the children’s expressive vocabulary knowledge was a significant predictor of their narrative comprehension in German (L1) but not in Swedish (L2, majority language). Lindgren and Bohnacker also analysed the different types of incorrect answers. They found that younger children (the 4-year-olds and 5-year-olds) tended to give more ‘I don’t know’ answers and provide nonsensical explanations than the older children (6-year-olds) who instead provided more logical (yet incorrect) explanations that are not part of the narrative context.59

Bohnacker, Öztekin and Lindgren (2020) studied a large group of Turkish-Swedish-speaking children in Sweden (N = 100, age 4–7). The children told all MAIN stories in the telling mode (two stories in each language). Bohnacker et al. found improvement of narrative comprehension with age for both languages, and a sharper increase for Swedish (L2) although no significant differences between Turkish and Swedish were found at group level. Similarly to Bohnacker and Lindgren (2021), Bohnacker et al. found that the accuracy rate patterns were similar across both languages on the individual comprehension questions and that the children performed better on Cat/Dog than on Baby Birds/Baby Goats. The authors drew the conclusion that there are subtle differences between the seemingly identical comprehension questions across the MAIN stories. Similarly to Lindgren (2018) and Lindgren and Bohnacker (2020), Bohnacker et al. found that the overall plotline question was the least accurately answered question in Cat/Dog (where the accuracy rates were in

---

59 An example of a logical (yet incorrect) explanation was found in the overall plot line question in Cat/Dog (D10: ‘Will the boy be friends with the cat/dog? Why?’) where a logical answer included the thought that the boy will be friends with the cat/dog because it is in fact the boy’s cat/dog). See Section 4.2.2 for more information on MAIN and Section 4.5.2.1 for the scoring of MAIN comprehension.
general high). For Baby Birds/Baby Goats, the question with the lowest accuracy rate was the question which probes the rationale of the internal state. Expressive vocabulary (CLT) in the respective language was found to significantly explain narrative comprehension scores.

Öztekin (2019), whose doctoral dissertation included the Turkish-Swedish-speaking children reported in Bohnacker et al. (2020), investigated the longitudinal development of 10 of these children from age 4 to 6. Öztekin’s findings in the longitudinal study generally confirm the results obtained in the cross-sectional study, namely that gains were stronger in the majority language Swedish than in the minority language Turkish.

In a study in Lebanon by Fiani, Henry, Prévost (2020), 48 simultaneous bilingual (Lebanese) Arabic-French-speaking children (age 4–9) were tested in both of their languages in the telling mode (Baby Birds, Baby Goats). The authors found age effects for both languages and no differences between the two languages. Questions targeting goals had in general a higher accuracy rate, particularly for the 4–5-year-olds and the 6–7-year-olds. Similar patterns in accuracy rates were observed for the individual questions across the languages. Results also showed that narrative comprehension scores were predicted by the children’s expressive vocabulary scores (ELO-L, Zebib et al., 2017) and exposure to storytelling.

In Finland, Kunnari and Välimaa (2020) compared the narrative comprehension abilities of Finnish-Swedish-speaking bilinguals (N = 16) with Finnish-speaking monolinguals (N = 16) all aged 5–6 in both the telling mode (Baby Birds, Baby Goats) and retelling mode (Cat, Dog). Results showed no difference between the language groups (monolinguals vs. bilinguals) or between the two languages of the bilingual children. Furthermore, the authors found no elicitation mode effect. Questions probing goals were answered with a higher accuracy rate than questions targeting internal states and their rationale.

Roch and Hržica (2020) studied the narrative comprehension skills of 30 Croatian-Italian-speaking sequential bilinguals (5–7) in both their languages (L1 Croatian, L2 Italian) using the telling mode (Baby Birds, Baby Goats). The children lived in Croatia and had been exposed to Italian (mean length 26 months). Results showed that children performed better in their L1 than in their L2 and that the accuracy rates for answering goal questions were higher than answering questions probing internal states. Furthermore, the authors found that receptive vocabulary (PPVT) correlated with L1 but not with L2 narrative comprehension scores.

In a longitudinal study by Gagarina, Topaj and Sürmeli (2020), a total of 57 Turkish-German-speaking children and Russian-German-speaking children were tested in their L2 (German) on three testing occasions (one year between each testing session). The children had a mean age of 3;6 (group 1,

---

60 Only 9 out of the 10 comprehension questions were asked by Roch and Hržica (2020).
When they were first tested. MAIN was elicited in the *telling* mode (Cat, Dog) and the *model story* mode (Baby Birds, Baby Goats). Results showed an age development for both groups and both narrative elicitation modes. Gagarina et al. found that children were able to answer the majority of the comprehension questions by age 5, after which no or little development was observed. The accuracy rate of answering *goal* questions was higher than *internal states* on all test occasions, for both groups and both elicitation modes. Answers to questions probing *theory of mind* were not fully developed even at the third testing.

To summarize, several studies that have used MAIN to investigate children’s narrative comprehension find an age development (at least till around age 6) (e.g. Bohnacker, 2016; Bohnacker & Lindgren, 2021; Fiani et al., 2020; Gagarina et al., 2020; Lindgren, 2018, 2019; Maviş et al., 2016; Roch et al., 2016;). In general, bilingual children tend to score similarly in both languages (e.g. Bohnacker, 2016; Bohnacker and Lindgren 2021; Bohnacker et al., 2020; Finai et al., 2020; Kapalková et al., 2016; Lindgren, 2018; Roch et al., 2016). As for the individual comprehension questions, a common finding seems to be that questions probing *goals* are answered with a higher accuracy rate than questions probing *internal states* (e.g. Bohnacker, 2016; Lindgren, 2018;) but see Kapalková et al. (2016). However, the children’s performance on narrative comprehension varies between the tasks even when narration is elicited in the same mode (*telling*) (Bohnacker et al., 2020; Bohnacker & Lindgren, 2021; Lindgren, 2018). Children score higher on Cat/Dog than on Baby Birds/Baby Goats. Comprehension of MAIN Cat/Dog thus appears to be easier. Some studies have found a relation between narrative comprehension performance and vocabulary production scores (Bohnacker et al., 2020; Fiani et al., 2020; Lindgren & Bohnacker, 2020) or with vocabulary comprehension scores (Roch and Hržica, 2020).

In the present study, the *telling* mode is used to elicit four MAIN narratives (two in each language) in Arabic-Swedish-speaking children in Sweden. The children are asked to answer each story’s respective comprehension questions. The narrative macrostructure comprehension results are reported in Section 7.1.3.

3.2.2 Narrative production with MAIN

The present study follows the *story grammar* framework that includes a *setting* (to specify *time* and *place*) and three episodes, as per the MAIN narrative elicitation tool (Gagarina et al., 2012, 2019). Each episode consists of causally and temporally related sequences that are based on a protagonist’s goal: an initiating event or *internal state* of the story character, the character’s *goal*, an

---

61 The children were only asked 9 out of the 10 comprehension questions.
an attempt or action to reach the goal, an outcome or consequence of the attempt in terms of the goal, and an internal state as a reaction to the outcome.

MAIN gives the opportunity to study macrostructure quantitatively by summing the number of story components produced in the whole narrative (story structure) and qualitatively by analysing the combination of produced episode components (story complexity).

Summarized below are studies that have examined narrative production skills using MAIN as an elicitation tool. The primary focus is on bilingual children, however, a few studies on monolingual children and monolingual adults are mentioned as well.

In a study by Gagarina (2016), 58 Russian-German-speaking children were tested for their narrative production skills in both of their languages using MAIN in the model story mode (i.e. hearing either Cat or Dog stories and then telling either Baby Birds or Baby Goats). The children were divided into three groups, preschoolers (N = 21, mean age 3;9), first graders (N = 15, mean age 7;0), and third graders (N = 22, mean age 9;3). For both languages, there was a sharp increase in the production of story components between the preschoolers and the first graders but no significant increase between the two elementary school groups. As for children’s story complexity, the preschool group was the only group that produced ‘no sequences’. The production of full episodes (GAOs) increased in both languages as the children became older (with a significant difference between the preschoolers and the first graders but not between the two elementary groups). A comparison between the languages showed only a difference in the oldest group where the production of GAOs in Russian (L1) was significantly higher than that in German (L2). Gagarina explains this difference via literacy training techniques usually done in Russian where children are taught explicitly about narrative episodic structure from an early age (since the first grade).

Kapalková et al. (2016) found that sequential bilingual Slovak-English-speaking children (age 5–6) scored higher on narrative production in their L1 (Slovak) than in their L2 (English). The expression of certain story components (e.g. initiating events and outcomes) was more frequent in L1 than in L2. Results showed that the children’s narrative comprehension skills exceeded their narrative production skills. Age development and story complexity were not examined.

For 16 Finnish-Swedish-speaking simultaneous bilingual children (5;0–6;6), Kunnari, Välimaa and Laukkanen-Nevala (2016) found no language effect on Finnish and Swedish narrative production story structure scores (for both telling mode and retelling mode). The children did, however, score significantly lower in the telling mode than in the retelling mode in Finnish but not in Swedish.62 As for episodic complexity, children performed similarly in

---

62 The narrative production scores of the bilingual children were compared to 16 Finnish-speaking monolingual same aged peers. The bilingual children’s Finnish narrative production scores
Finnish and Swedish. Elicitation mode (telling vs. retelling) had an effect on story complexity with more produced episodes in the retelling mode. Age effects were not analysed.\(^{63}\)

Roch et al. (2016) found significant effects of age, language and mode (telling vs. retelling) on the narrative production of Italian-English-speaking sequential bilinguals. The younger children (5–6-year-olds) scored lower than the older children (6–7-year-olds) in both languages. There were no differences between the languages for the older age group, but the younger children provided more story components in their L1 (Italian) than in L2 (English). The children provided more macrostructure components in the retelling mode than in the telling mode. As for macrostructural story complexity, age effects were found. Similarly to the production of story components, the youngest age group produced more complex episodes in L1 than in L2, whereas no such language differences were found for the older age group.

Bohnacker (2016) found an effect of age for the narrative production scores of 5-year-old versus 6–7-year-old English-Swedish-speaking bilinguals. No language effects were found. For both age groups, the most commonly expressed story components were attempts and outcomes. Goals, internal states, and setting (time and place) were expressed to a much lesser degree. The production of complete episodes (GAO sequences) was infrequent in both age groups and for both languages, with a slight increase from 5 to 6–7-years. In both languages, the production of story components lagged behind their comprehension. Goals and internal states were understood to a much larger greater degree when queried via comprehension questions that they were spontaneously expressed in narration.

In Lindgren’s (2018) study on narrative production of 72 Swedish-speaking monolinguals, 46 German-Swedish-speaking bilinguals and 48 Turkish-Swedish-speaking bilinguals, age development was found in all language groups in Swedish (the majority language).\(^ {64}\) For both bilingual groups, age effects were observed for Baby Birds/Baby Goats scores but not for Cat/Dog scores in the respective home language (German or Turkish). The Turkish-Swedish-speaking children scored equivalently in their two languages, whereas the German-Swedish-speaking children scored higher on Swedish than on German. In all languages, attempts and outcomes were expressed to a larger degree than other types of story components. Furthermore, the production of attempts and outcomes increased with age, but not the production of goals. As

\(^{63}\) Results showed that monolinguals scored significantly higher on Finnish story components but not on story complexity than the bilingual children.

\(^{64}\) In a follow-up study on a sub-group of the monolingual Swedish-speaking children (N = 17), Lindgren (2019) found a steep increase in structure scores from age 4;4 to age 5;10 but a minor development from 5;10 to 7;4, indicating the reaching of a plateau in the development of narrative macrostructure around age 6. Lindgren also found a story effect where children scored higher on Baby Goats than on Baby Birds.
for story complexity, the German-Swedish-speaking children produced a higher proportion of sequences than the Turkish-Swedish-speaking children in both Swedish and the respective home language. However, the production of full sequences (goal-attempt-outcome, GAO) in both languages did not differ between the two bilingual groups. Age development for the production of GAOs was only found in the majority language Swedish but not in the home language German or Turkish. For no group or language was there any task effect between Cat/Dog and Baby Birds/Baby Goats in macrostructure production. As for the difference between narrative production and narrative comprehension, results showed that children scored significantly higher in comprehension than in production. Furthermore, in all three languages and in both tasks, children understood a higher number of goals than what they produced.

Lindgren and Bohnacker (2022) argue the fact that the German-Swedish-speaking children (in Lindgren, 2018) scored higher on narrative production in Swedish than in German goes against the assumption that the production of story structure is invariant across a bilingual child’s two languages as found by various studies (e.g. Akinci et al., 2001; Bohnacker, 2016; Fiestas & Pena, 2004; Pearson, 2002). Furthermore, Lindgren and Bohnacker found that expressive vocabulary (CLT) scores predicted the children’s narrative production in German but not in Swedish.65

Bohnacker, Lindgren and Öztekin (2022) studied the narrative production skills of 100 Turkish-Swedish-speaking children (4–7) who were administered MAIN in the telling mode. Age development was found for both of the children’s languages, more prominent for Swedish than for Turkish. No language differences were found. There was no task effect, the children scored similarly on Cat/Dog and Baby Birds/Baby Goats. The most frequently produced story components in both languages were attempts and outcomes, and the least commonly produced components were internal state terms as reaction and setting. As for story complexity, in both Cat/Dog and Baby Birds/Baby Goats, the production of ‘no sequence’ was most common in both languages (but more so in Turkish than in Swedish) and in all age groups but it decreased with age. The second most frequently produced sequence was attempt-outcome (AO). Full episodes (goal-attempt-outcome, GAO) were generally rare in both languages. For both Cat/Dog and Baby Birds/Baby Goats in both languages, the production of GAO increased with age but was still quite rare even for the oldest age group.66

65 The effect of narrative length, language input and the effect of language of the first testing were also analysed, but will not be reported here.
66 Bohnacker et al. (2020) also studied the effect of narrative length, expressive vocabulary (CLT), length of L2 Swedish language exposure and estimated daily language input on narrative production scores. Results were similar in both languages and showed that children who
Öztekin (2019), whose doctoral dissertation included the Turkish-Swedish-speaking children reported in Bohnacker et al. (2022), also investigated the longitudinal development of 10 of these children from age 4 to 6, in both languages. Öztekin found that narrative production increased for both Turkish and Swedish but that gains were stronger for majority language Swedish. This result confirms the general pattern of development with age in the cross-sectional study.

In Lebanon, Fiani, Henry and Prévost (2022) examined the narrative production of 69 (Lebanese) Arabic-French-speaking children (age 4;3–9;9) using Baby Birds/Baby Goats in the telling mode. The children were divided into three age groups: 4–5-year-olds, 6–7-year-olds and 8–9-year-olds. For the production of story components, no language differences were found between the age groups, irrespective of language dominance. The fact that the children scored similarly in both Lebanese Arabic and French could be related to the Lebanese setting which includes more than one majority language (see Section 1.3 for an overview on the Lebanese bilingual/multilingual setting). The most frequently produced story components in both languages were attempts and outcomes whereas the production of goals was in general much lower. The production of story complexity also increased with age, with a sharp increase between the 4–5-year-olds and the 6–7-year-olds. The youngest age group did not manage to produce any complete sequence (GAO), and was the only age group that produced ‘no sequence’ (67% in Lebanese Arabic and 60% in French). GAOs were produced by the older two age groups and a noticeable increase in the production of GAOs was observed from the 6–7-year-olds to the 8–9-year-olds in both languages.

Fichman, Walters, Armon-Lotem, and Altman (2022) examined the difference in narrative production skills (in the retelling mode for Baby Birds and Baby Goats) between two groups of Russian-Hebrew-speaking bilingual children (5;7–6;7) where one group was L1 Russian-dominant (N = 19) and the second group was L2 Hebrew-dominant (N = 19). Results showed that the two groups performed similarly in the production of story components and story complexity. Furthermore, in both groups, there was no significant difference between the two languages. The most produced story components were attempts and outcomes whereas goals were expressed less frequently. GAOs were produced to different degrees across the three episodes, but most frequently expressed in the second episode (the cat/fox wants to catch the birds/goat, climbs/jumps, catches its leg). The authors conclude that language dominance did not impact the production of narrative macrostructure. However, as mentioned by Bohnacker and Gagarina (2022), another reason behind the lack of any language dominance effect might be related to Fichman et al.’s had longer narratives and children whose expressive vocabulary scores were high scored generally high on narrative production, whereas children’s language exposure and daily language input did not measurably affect their narrative production.
narrative elicitation method, the *retelling* mode, where the children first listened to the story and were then asked to retell it.

In Tribushinina, Irmawati and Mak’s (2022) study on 32 Indonesian-Dutch-speaking children (5;0–11;9) in the Netherlands, the children told Cat/Dog stories first in Indonesian (L1) directly followed by Baby Birds/Baby Goats in Dutch. Results showed that story structure scores in L1 were positively related to age, whereas L2 scores were not. Age was not a predictor of story complexity in either language. For both story structure and story complexity, performance in L1 was much better than performance in L2. The authors suggest that these results were obtained since the majority of the children had short exposure to L2 Dutch. Another reason why the children performed better in L1 than in L2 could be related to the testing methodology. The two language testing sessions were administered by one experimenter on the same day which might have encouraged code-switching and led to fatigue. In the present study, two different native (or near-native)-speaking experimenters tested the children on two separate occasions.

Even though *child* narration is the topic of the present study, it is worth mentioning the only study that has been published on *adults* using MAIN, in order to have an insight into what ‘adult-like’ performance looks like.

Gagarina, Bohnacker and Lindgren (2019) found that adults (age 19–41, monolinguals speaking German *N* = 30, Swedish *N* = 19, or Russian *N* = 20) produced different macrostructural components (in Baby Birds/Baby Goats) to similar degrees, irrespective of what their mother tongue was. *Goals* were produced to a lesser degree than *attempts* and *outcomes* that were each produced at least 80% of the time irrespective of episode and language. *Goals* were produced at around 50% in the first two episodes and at round 35% in the third episode. The mean score for total story structure on MAIN was 11–12 points out of maximum 17 points. Hence, the authors suggest that if a child scores 11 or 12 (out of 17) on narrative story production, this result can be considered ‘adult-like’. Furthermore, adults scored significantly higher in Baby Goats than in Baby Birds. As for story complexity, 40% of all the episodes produced (by each language group) contained a full episode (GAO-sequence). The most frequently produced sequence was attempt-outcome (AO) (German: 42%, Russian: 43%, Swedish: 49%). ‘No sequence’ was produced far more rarely (German: 9%, Russian: 10%, Swedish: 7%). The vast majority of adult narrators produced at least one GAO per story (German: 77%, Russian 70%, Swedish 74%) and thus reached the highest level of story complexity. There were no signs of differences between the languages. In short, performance across the languages was very similar, but differed somewhat from one episode to another.

To sum up, studies that have used MAIN to test children’s narrative production found age effects in both languages of bilingual children but generally no language effects, meaning that bilingual children (as a group) scored similarly in both of their languages (e.g. Gagarina, 2016; Roch et al., 2016;
Bohnacker, 2016; Lindgren, 2018; Fiani et al., 2022; Bohnacker et al., 2022; Lindgren and Bohnacker, 2022). In general, no differences were found between the different tasks (Cat/Dog and Baby Birds/Baby Goats) (e.g. Lindgren, 2018; Bohnacker et al., 2022) which is different than for narrative comprehension where differences between tasks have been found (see Section 3.2.1). Only few studies found story differences in the children’s (and adults’) performance on Baby Birds and Baby Goats (Lindgren, 2019, Gagarina et al., 2019). A general finding is that different story component types are produced at different frequencies. Attempts and outcomes seem to be produced to a much higher degree than goals and internal states. In general, studies that have compared narrative comprehension scores with narrative production scores found that children scored much higher in narrative comprehension than in production (e.g. Kapalková et al. 2016; Lindgren, 2018; Öztekin, 2019). For story complexity, the production of ‘no sequence’ appears to be common among younger children (but decreases with age), whereas the production of full episodes (GAOs) is very rare at age 4-5 but becomes more frequent as children grow older (e.g. Bohnacker, 2016; Bohnacker et al., 2022; Fiani et al., 2022; Gagarina, 2016). It also seems to be the case that macrostructure in production is not fully developed yet by age 7 and that many children have some way to go till they use adult-like production (Gagarina et al., 2019).

As previously mentioned, the children in the present study were tested in the telling mode using MAIN as an elicitation tool. The Arabic-Swedish-speaking children told two narratives in both of their languages. Results for narrative macrostructure production are reported in Section 7.2.3.
4. Methods

This chapter reports the methods used in the three studies that are included in the present dissertation: the cross-sectional study in Sweden, the longitudinal study in Sweden and the cross-sectional study in Lebanon. Section 4.1 describes the procedures for the recruitment of participants and Section 4.2 describes the testing materials used. Section 4.3 outlines the data collection procedure and Section 4.4 explains the transcription method used for the lexical and the narrative data. Section 4.5 describes the scoring procedures of the lexical and the narrative responses. Finally, Section 4.6 includes ethics considerations concerning the three studies. Statistical procedures are not described in the present chapter, but rather in the results chapters for the lexical and narrative tasks (Chapters 6, 7, 8, 9).

4.1 Recruitment

Section 4.1 describes the recruitment process in the cross-sectional study in Sweden (4.1.1), the longitudinal study in Sweden (4.1.2) and the cross-sectional study in Lebanon (4.1.3). Each section also outlines how background information about the participants was gathered and how the children’s anonymity was ensured.

4.1.1 Recruitment for the cross-sectional study in Sweden

Recruitment for the cross-sectional study in Sweden started in May 2017 and data collection took place between September 2017 and March 2019. The criteria for recruitment were that the children were between ages 4 and 7, were able to speak (to some degree) both Arabic and Swedish and did not have a known diagnosis of Developmental Language Disorder (DLD). The first criterion, being aged between 4 and 7, was chosen in order to study children’s language development from preschool age transitioning into school age (where formal education begins). Concerning the second criterion, to be an active language user of Arabic and Swedish, children were not expected to have near monolingual competence in both languages, but rather to be able to at least produce meaningful utterances in both languages. This criterion was important in order to be able to analyse and compare narrative macrostructure
production abilities in both languages. Hence, children who only had receptive knowledge in one of the languages were excluded from the study. The third criterion, not to have a DLD-diagnosis, was agreed upon in the BiLI-TAS project since one of the overarching project’s aims (but not an aim of the present study) was to explore whether DLD-diagnosed children performed similarly or differently on certain language tasks in comparisons with children with typical language development. (For a comparison of Arabic-Swedish-speaking children with and without a DLD diagnosis, see Öberg (2020).) One of the aims of the present study was to statistically compare the effect of background factors on vocabulary scores. Because DLD-diagnosed children have been found to have smaller vocabularies compared to typically developing bilingual peers (e.g. Khoury Aouad Saliby et al., 2017) and often have general language comprehension difficulties (e.g. Bishop et al., 2017), including them in the present study might have masked effects of different language input factors on vocabulary and their general performance on the narrative task (for example, having difficulty in understanding the task requirements or the comprehension questions).

The aim was to include 100 children who were evenly distributed across the ages 4, 5, 6 and 7. One hundred was considered a manageable number of participants for a small research team to recruit, test and analyse, yet also a large enough number of children in order to perform meaningful statistical analyses on and to represent the targeted bilingual group.

The recruitment process initially targeted Arabic-Swedish-speaking children regardless of which Arabic variety the children spoke. However, early on in the data collection process this was restricted to include foremost speakers of Levant and Iraqi varieties. This decision was made to ensure mutual understanding between experimenter and child during the Arabic testing session.

Children were recruited from urban areas in Eastern Central Sweden, more specifically from the greater Uppsala, Stockholm and Västerås regions. Sampling throughout more regions in Sweden was not possible due to limited human resources, as well as time and budget constraints. Furthermore, in Sweden, random sampling of Arabic-Swedish-speaking individuals per se is not possible since no statistic information is gathered by Statistics Sweden on what language(s) individuals speak.

67 For information on the Arabic varieties spoken by the participating children (as well as the excluded ones) in the cross-sectional study in Sweden, refer to Section 5.1.1.
68 Statistics Sweden, or Statistiska Centralbyrån in Swedish, is a government agency that gathers statistics on a national level. Information from Statistics Sweden is available in regard to which country individuals were born in and which country they had emigrated from. However, knowing that a person was born in or has emigrated from an Arab country does not guarantee that the person is a proficient Arabic speaker or that her/his child is. For instance, there are speakers of minority languages such as Aramaic/Syriac, Armenian and Kurdish who have migrated from Arab countries and have chosen to teach their children the minority language and/or Swedish (rather than Arabic).
The children participating in the cross-sectional study in Sweden were reached by contacting 31 schools and preschools as well as by contacting 15 associations and congregations where Arabic is a potential common spoken language. Additionally, children were also recruited by contacting Arabic-speaking MTI teachers and through Arabic-speaking research assistants who recruited via their personal connections.

All parents were handed an informative leaflet, an informed consent form and a questionnaire and were asked to return the latter two forms filled, if they were interested in participating. The filled-in forms were gathered prior to meeting with the child. In the few cases where this was not possible, the parents were asked to hand in the forms as soon as possible. The gathering of the informed consent and questionnaire was usually done via the help of (pre)school staff or group leaders at the associations and congregations.

The informative leaflet (available in both Arabic and Swedish) contained information about the study and its objectives, the inclusion criteria for participation and contact details of the author and other members of the research team and research assistants. The parents were reminded about their possibility to withdraw their participation at any time without providing justification. Also, an Arabic-Swedish video was recorded to ‘give a face’ to some of the experimenters whose names were on the informative leaflet. The author, along with another PhD student (Linnéa Öberg) made a short bilingual video to give the parents a concrete idea of the testing procedure and the elicitation materials used in the testing sessions with their child.

The parental questionnaire queried information on the linguistic and social background of the children and their parents. The parents could answer the questions in the language of their choice, in either Arabic or Swedish. The

---

69 Recruitment started by contacting (pre)schools; later, recruitment shifted towards associations and congregations where members are potential speakers of both Arabic and Swedish. The main Arabic-speaking experimenter (the author) visited each organization’s head leader and explained in detail the importance of the current project, showed the test materials and illustrated how a meeting with a child would look like. Once approval was obtained from the organization’s head leader, group leaders for the relevant children’s age groups were contacted and were asked to distribute the information leaflets and consent forms to families of children who fulfilled the criteria. Recruiting via associations and congregations proved to be fruitful as it gave a quicker positive acceptance rate amongst the parents who had the opportunity to meet the experimenter in person and put questions face to face.

70 This video was made when the recruitment procedure shifted towards targeting congregations and associations. The video was often sent (via email or telephone) to the group leader to spread it further to potentially interested parents whose children were within the targeted age range and fulfilled the recruitment criteria. Some parents, especially those whose children were recruited via the congregations and associations, held conversations with the author where they were provided with oral information concerning the aims of the present study, in addition to being handed written information.

71 The parental questionnaire was developed by the BiLI-TAS project in 2014-2016. The questionnaire is available in several languages, including Arabic and Swedish. Prototypes of the questionnaire were discussed with native-speaker linguists, minority community members and speech-language pathologists. These prototypes were later piloted, and finalized after some changes.
questionnaire for the cross-sectional study included 36 questions targeting (amongst other aspects):

- The child’s age of onset in each language
- The child’s estimated daily language exposure
- The child’s language proficiency as rated by parents
- Languages spoken at home and outside the home
- Information regarding the parents’ languages and education
- Concerns regarding the child’s language development\(^\text{72}\)
- Information regarding Mother Tongue Instruction attendance and exposure to other Arabic varieties
- Additional language exposure through home activities (such as storytelling and shared book reading).

In the consent form, apart from providing informed consent, the parents were asked to specify which Arabic variety the child spoke. This information was particularly important for the preparation of the meeting with the child in the Arabic testing session. The form also included clear information that the testing sessions would be video and audio recorded and that the anonymity of the child and the parents would be preserved.

All the children who participated in the cross-sectional study in Sweden received an individual code to preserve their anonymity in further data handling. Each code started with the five letters ‘BiAra’ (short for Bilingual Arabic), followed by one digit representing the child’s age group (4, 5, 6, or 7), and a unique two-digit number. To illustrate, the unique code ‘BiAra4-24’ represents a 4-year-old Arabic-Swedish-speaking bilingual child in Sweden.

### 4.1.2 Recruitment for the longitudinal study in Sweden

Recruitment for the longitudinal study started in November 2019 and data collection took place from November 2019 to February 2021. The aim was to include at least 10 children who were 4 years old when they first participated in the cross-sectional study and retest them when they were 6. Nineteen families were re-contacted out of 24 families that took part in the cross-sectional study.\(^\text{73}\) The parents were contacted by the author and were asked whether the family was interested in participating in a follow-up study. Upon oral consent, the parents were asked to sign a new informed consent and to fill in a questionnaire. Because of the COVID-19 pandemic, only 10 children could be re-tested (with immense recruitment and rescheduling efforts). The original aim

---

\(^{72}\) Questions regarding language development were added to ensure that none of the participants had a diagnosis of Developmental Language Disorder (DLD, also known as Language Impairment, LI).

\(^{73}\) Five families from the 4-year-old group in the cross-sectional study could not be re-contacted due to Swedish national restrictions during the COVID-19 pandemic.
was to test the children 2 years (approximately 24 months) after the first testing session. However, due to contact restrictions during the pandemic, this was not always possible, see Section 4.3.3.

The questionnaire for the longitudinal study, available in Arabic and Swedish, included a reduced number of questions (selected 17 questions) from the cross-sectional questionnaire, eliciting information about language exposure and use that might have altered since the first testing, approximately 2 years earlier. The questions queried which languages were spoken at home, the language(s) spoken by the school staff, language-stimulating activities at home, the parents’ evaluation of the child’s language skills, and their estimation of the child’s language input throughout the day.

Additional background information was gathered via interviews with the parents and with the child. Oral interviews with the parents were held by the author (a native speaker of Arabic) in Arabic in an informal tone. These interviews were not recorded but rather notes were taken by the interviewer to keep the conversation casual. Some of the interviews were held at home, while others were conducted over the phone. The parents could choose whether one or both of them wanted to participate in the interview. The interview questions targeted language use in the home and whether this had changed over time, the child’s code-switching behaviour and the parents’ strategies towards this behaviour, whether the child had friends or teachers who spoke Arabic with them, whether the parents were content with the child’s language use in the family, and whether the family read any Arabic or Swedish books at home. Further questions included whether the child was exposed to additional languages (other than Arabic and Swedish) and whether there were any potential concerns regarding the child’s or anyone else in the family’s language development.

An informal oral interview with the child was conducted as part of the warming up during the Swedish testing session. The interviewer (who was also the experimenter of that session) was either Linnéa Öberg, a native speaker of Swedish and a PhD student in linguistics, or Ute Bohnacker, a near-native speaker of Swedish and professor of linguistics. The children were asked which languages they knew, which languages were spoken at home, their preferred language, whether they spoke the home language outside the home, and whether the children watched any movies or did any (joint) book reading in Arabic or Swedish at home or (pre)school.

The children who participated in the longitudinal study are again referred to by code, to preserve their anonymity. They kept the same code they had in the cross-sectional part, with the words ‘Follow Up’ added. To illustrate, the child whose code was ‘BiAra4-24’ (when participating in the cross-sectional study), was given the code ‘BiAra4-24 Follow up’ to represent the same child tested approximately 2 years later.
4.1.3 Recruitment for the cross-sectional study in Lebanon

Recruitment preparations for the cross-sectional study in Lebanon started in Sweden prior to traveling to Lebanon for the data collection. The author contacted 11 preschools and schools and 5 congregations and associations via email or telephone to inform them of the ongoing studies and ask their collaboration in recruiting bilingual children in Lebanon. The children were recruited through personal contacts and with assistance from acquaintances of the author. The contacted schools, congregations and associations were located in greater Beirut, Maten, Kiserwan, Byblos and Batroun districts.

Recruitment (and data collection) for the cross-sectional study in Lebanon took place in August 2018 and March 2019. The criteria for recruitment were that the child was bilingual, spoke Arabic plus a second language (usually English or French), and was not diagnosed with DLD (thus supposedly having a typical language development). The aim was to recruit an equivalent number of children to that of the cross-sectional study in Sweden ($N=100$), for comparative and statistical purposes.

Similarly to the cross-sectional study in Sweden, the parents in Lebanon were handed an informative leaflet and were asked to sign an informed consent in addition to filling in a questionnaire if they wished their child to participate. The questionnaire consisted of the same 36 questions as the cross-sectional study in Sweden; however, the questions were slightly adjusted to fit the Lebanese context (for example, Swedish as a second language was replaced by English or French).

To preserve anonymity, each child in the Lebanese cross-sectional study received a unique code, starting with the letters BiLeb (short for Bilingual Lebanese), followed by a digit corresponding to the child’s age group (4, 5, 6 or 7) and a unique two-digit code. For instance, ‘BiLeb7-07’ is a bilingual 7-year-old child who participated in the cross-sectional study in Lebanon and who speaks Arabic as well as English or French.

4.1.4 Summary

The process of recruiting and collecting data from all the children participating in the present study stretched over 3 years and 10 months (from May 2017 to February 2021). The children were recruited mainly through (pre)schools and

---

74 This recruitment procedure was very efficient and the author was able to test the desired number of children in a relatively short time. What was not possible, however, was to recruit children from public schools, since special permission would have to be sought at the Ministry of Education and Higher Education (MEHE). This procedure was judged to be time consuming and hence dropped. It so happened that the children that were recruited via private initiatives and with the help of contacts all attended private schools.

75 In Lebanon, these were handed out in Arabic (MSA) only. Again, both the informative leaflet and the informed consent contained clear information regarding audio and video recording of the testing sessions and the possibility to withdraw their participation at any point.
organizations. All parents received an informative leaflet, an informed consent form and a questionnaire. Additionally, interviews were held with the families who participated in the longitudinal study. Table 4.1 provides a summary of the different methods used to gather background information in the three studies. For a description of the participants, see Chapter 5.

Table 4.1. Methods for gathering background information.

<table>
<thead>
<tr>
<th>Cross-sectional study in Sweden</th>
<th>Longitudinal study in Sweden</th>
<th>Cross-sectional study in Lebanon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental questionnaire (36 questions)</td>
<td>Parental questionnaire (17 questions) Informal parental interview (18 questions) Informal child interview (17 questions)</td>
<td>Parental questionnaire (36 questions)</td>
</tr>
</tbody>
</table>

4.2 Materials

This section provides information about the elicitation materials used in the present dissertation as well as the procedures for adapting the materials to the language background of the Arabic-speaking participants.

In order to test the vocabulary skills of the children, the Cross-linguistic Lexical Task (CLT; Haman et al., 2015) was used. For the assessment of their narrative skills, the Multilingual Assessment Instrument for Narratives (MAIN - Revised, Gagarina et al., 2019) was used. The test battery also included four non-word repetition (NWR) tasks to test the children’s phonological processing. These results are not reported in the present study, but see Öberg (2020).

The CLT and MAIN were administered in the cross-sectional and longitudinal studies in Sweden in both of the children’s languages. As for the cross-sectional study in Lebanon, only the (Lebanese) Arabic CLT was used. MAIN narratives are also available for 100 children in Lebanese Arabic. Due to time constraints, MAIN data from Lebanon will not be analysed in the present dissertation.

4.2.1 Cross-linguistic Lexical Task (CLT)

The Cross-linguistic Lexical Task (CLT; Haman et al., 2015) is a picture-based vocabulary test that consists of four parts: noun comprehension, verb comprehension, noun production and verb production. Each part contains 30 target words (plus 2 practice words) summing up to a total of 120 words (plus 8 practice words).
The CLT was developed within the COST Action IS0804 (2009–2013) by researchers in Working Group 3 (Lexical and Phonological Processing). The aim was to design a comparable lexical tool suitable for assessing comprehension and production of nouns and verbs in bilingual children as well as children with Developmental Language Disorder (DLD). The CLT was designed in a manner so that a typically developing monolingual 5-year-old child would achieve relatively high results on the tasks. The validity of the CLT was confirmed in a large-scale study with 17 languages (see Haman et al., 2017).

CLT measures both receptive and expressive vocabulary knowledge using the methods of picture naming (production) and picture selection (comprehension) since they are believed to be tasks least involving other types of linguistic or conceptual skills (Haman et al., 2015). The production part consists of 60 target words (30 nouns and 30 verbs) plus 4 practice words (2 in each section). The child is asked to name the object or action that is shown in the picture by answering the experimenter’s prompt question what is this? (noun production) and what is s/he doing? or what is happening? (verb production). See Figure 4.1 for examples.

**Figure 4.1.** Small-scale examples from the Arabic CLT production, nouns (left) and verbs (right). ©University of Warsaw

**Figure 4.2.** Small-scale examples from the Arabic CLT comprehension, nouns (left) and verbs (right). ©University of Warsaw.

---

76 For more information regarding COST Action IS0804 and the CLT, see Armon-Lotem, de Jong & Meir (2015) and https://www.bi-sli.org/.
The comprehension part consists of 60 target words (30 nouns and 30 verbs) plus 4 practice words (2 in each section). The child is asked to identify the correct corresponding picture out of 4 presented pictures (each target picture has 3 distractors). The target word is embedded in the prompt question, for example, where is the pear? (noun comprehension) or who is squeezing? (verb comprehension). The child is asked to identify the correct target picture by either pointing at the picture or naming the number assigned to it. See Figure 4.2 for examples.

The CLT subtasks are presented to the child in booklets of laminated pages (a total of four different booklets, one for each subtask). The pictures are printed with coloured ink on white paper in A5 (production) or A4 (comprehension) format.

CLT versions currently exist for 34 different languages. These different CLT language versions are not identical in terms of target words (and distractor items), however all the target words were chosen from one common word list of basic concepts for objects (158 nouns) and actions (142 verbs), and their corresponding pictures from a picture base of 1024 pictures. The word list contains words that are basic, concrete, everyday words with high imageability. Pictures that correspond to the target concepts have been balanced regarding the number of female and male characters while avoiding gender and racial stereotypes. All pictures are homogeneously coloured drawings with no shading and no background.

The different CLT language versions are, thus, not translated from one language version, but rather constructed based on common standardized principles (see Haman et al., 2015). For each CLT language version, target words have been selected according to that language’s cultural and linguistic context as well as each item’s difficulty. The item’s difficulty is based on an estimation of that item’s age of acquisition and its phonological and morphological complexity. See Haman et al., 2015 for a detailed description of the selection of the target items and how CLT was constructed. Hence, each language version has its own unique combination of target words. Although no two language versions are identical, there may still be some overlap in the target items between the versions, since as previously mentioned, all the versions were created from the same list of concepts. The different CLT language versions are believed to be comparable across the languages.

The language versions used in the present study are Swedish, Lebanese Arabic and adapted versions of Lebanese Arabic. The Swedish CLT version was developed by Gisela Håkansson, Natasha Ringblom and Josefin Lindgren (Ringblom, Håkansson, & Lindgren, 2014) and the Lebanese Arabic CLT version was developed by Christel Khoury Aouad Saliby, Edith Kouba Hreich

---

77 For some language versions, digital formats are also available.
78 New language versions are continuously being prepared, see website: https://multilada.pl/en/projects/clt.
and Camille Messarra (Khoury Aouad Saliby, Kouba Hreich, & Messarra, 2017). The adapted versions of the Lebanese Arabic CLT were developed by the author (Haddad, 2017) as described below. There are no cognates in the Arabic and Swedish CLT language versions.

4.2.1.1 Adaptation of the CLT

Since the Arabic-speaking population in Sweden originates from several countries and thus speaks different Arabic varieties, the Lebanese Arabic CLT could not be straightforwardly used with children in Sweden as not many of them speak Lebanese Arabic. Therefore, the Lebanese version needed to be adapted to Arabic colloquial varieties that are most relevant in the Swedish context, namely Syrian, Palestinian, Iraqi as well as to Modern Standard Arabic (MSA). The actual target responses and target words (as well as their corresponding pictures), were kept unchanged from the Lebanese Arabic version, but the lexical prompts and accepted answers were revised.

Adaptations needed to be made for both the comprehension and production tasks to avoid miscommunication and ensure mutual understanding between the child and the experimenter. In the comprehension tasks, it was crucial for the child to hear a word that s/he could actually be familiar with (i.e. a word that is used in her/his Arabic variety), if not, the child would not be able to identify the correct test item. In the production tasks, if a child responded by using a correct label in her/his Arabic variety, but if this word was not part of the accepted target word list, then the answer would be unjustly scored as incorrect. Consequently, adaptations were developed by the author for both the lexical comprehension and production tasks. The aim was to ensure that a child, regardless of which Arabic variety s/he spoke, was tested and scored fairly when examined on her/his Arabic vocabulary knowledge.

For the CLT production tasks, the prompt questions (what is this? and what is s/he doing? / what is happening?) were added in different Arabic variety instruction sheets. Moreover, the (Lebanese) production target word lists were complemented by additional colloquial synonyms to be considered as correct when scoring the production tasks. For the CLT comprehension task, new prompt questions (where is the...? and who is...?) and test items were constructed for all nouns and verbs in the different varieties, so as not to disadvantage the children by asking them about a word in a variety they do not understand or use.

The adaptation process started by consulting Arabic (MSA)-English as well as Arabic dictionaries targeting the colloquial varieties. Also, the author consulted native-speaker informants (N = 9, ages 30–45 years) who spoke varieties of Arabic from nine different locations in the Levant and Iraq. The informants were sent lists of all the target concepts (for both comprehension and

79 See Section 4.5.1 for a detailed description of the scoring procedure of the CLTs.
production) in English, Swedish or Lebanese Arabic, and were asked to express the same concepts in their own Arabic variety.⁸⁰ The author also consulted native Arabic-speaking Semitic scholars (N = 7, ages 25–45) to name all the target concepts in their own variety. The Semitists were handed pictures of the target items (both comprehension and production) and were asked to audio record the lexical labels for the pictures.⁸¹ Parents of participating children (N = 6) were also consulted (on the telephone or in person) when unfamiliar/unclear utterances occurred regarding their own children’s responses or the responses of other children who were reported to speak a similar Arabic variety to their own.

It was important to complement the (Lebanese) Arabic list of lexical labels from a variety of sources (dictionaries, native informants, Semitic scholars and parents) so as to obtain a large comprehensive list with possible synonyms suitable for Arabic-speaking children in Sweden. As previously mentioned, some Arabic varieties differ tremendously from one region to another within the same country. This issue was handled, to the best of the author’s ability, by contacting as many reliable informants from different regions within the same countries. It is inevitable though that there are missed regions/colloquial variants and unrepresented minorities. See discussion in Chapter 11. In Syria, data was mostly gathered from informants who were native speakers of Arabic varieties spoken in greater Damascus, Homs province and Jazira province. In Palestine, informants mostly spoke the Arabic dialects of greater Jerusalem, Ghaza strip, and north-west Palestine. In Iraq, the informants mostly spoke Arabic varieties of greater Baghdad, Mosul region, and Najaf region.

By the end of the adaptation procedure, one CLT test item could have up to five different lexical choices (for the Arabic varieties combined). For example, the test item ‘bed’ had five possible target words (taḥet, serir, farše, šurpaye, qaryole) and the test item ‘nose’ had four (anf, munḥar, ḥaršum, ḥa-shim).⁸² However, two or three lexical options (synonyms) for each target

---

⁸⁰ If a target word was unclear, the author resolved the issue by sending a picture of the word and asked the native speaker for the target word based on the picture. Sending out all the pictures to the native speakers was not always possible due to practical reasons.

⁸¹ The Semitist scholars were met in person by the author at two Arabic linguistics conferences. The Semitists were handed small-scale versions of the CLT pictures that the children see in a test session. For the production part, the Semitists were asked to name the item/concepts. In the comprehension part, the target word was pre-identified (circled on the paper containing the 4 pictures). The Semitic language experts were asked to name those specific items/concepts in their native Arabic variety.

⁸² Some answers were provided by the informants as ‘correct in their own variety’. These answers, although not always in agreement with the dictionaries for that corresponding Arabic variety, or with the MSA meaning of the word, were still taken into consideration (but were not the first word provided to the child). See below (Section 4.2.1.2) on the administration of the Arabic CLT. Other versions of the target words in the CLT lists were also mentioned by the informants, for example, the plural form (manaxir) and phonological alterations (xashmi or xshaam). Subtle phonological alterations were not always noted down in the CLT comprehension prompt sheets (used by the experimenter), but were often produced by the experimenter at the time of data collection.
word were more common. The different lexical forms were incorporated into the prompts for the CLT Arabic comprehension and were added as correct synonyms to the target answers for CLT Arabic production. Consequently, the existing Lebanese Arabic CLT version (Khoury Aouad Saliby et al., 2017) was enriched with lexical options, thus creating further Arabic versions, namely Syrian, Palestinian, Iraqi (Baghdad variety and Mosul variety).

4.2.1.2 Administration of the CLT

The CLT was administered according to the standard procedure as indicated by Haman et al. (2015). In the production part, the child was shown one picture at a time and asked to name the object or action presented. If the child answered in a language different from the target language (for example, if the child answered in Swedish in the Arabic session) the experimenter would repeat the question once more emphasizing that she was interested in the answer in the target language, or simply say ‘and in Arabic/Swedish?’ If the child did not answer or ask for clarification, the experimenter would repeat the question once more before proceeding to the next item. No gestural, semantic or phonological cues were provided to help the child.

In the comprehension part, the child was shown a page with four pictures at a time and was asked to identify the correct target picture. If the child did not respond or ask for clarification, the experimenter would repeat the question once more before continuing to the next page. When conducting the comprehension subtasks in Arabic in the cross-sectional and follow-up studies in Sweden, the testing procedure was slightly modified from the standard procedures in order not to disadvantage children that were unfamiliar with the colloquial test item provided.

As previously mentioned, the families who participated in Sweden originated from different Arabic regions, therefore the children spoke different Arabic varieties. Furthermore, many children were reported to have had contact with other Arabic speakers whose variety was different from their own, for example, through peers or school staff. (See Chapter 5, on participants.) Hence, for certain concepts, a child may be familiar with some lexical form more than others. After piloting and further discussion with the BiLI-TAS research team,83 it was decided to give the Arabic-speaking children in Sweden more options of synonyms for the comprehension part if they appeared not to understand the lexical items used by the experimenter. Prior to the Arabic test session, the experimenter was informed which Arabic variety that the child spoke (according to the parents via the informed consent form and/or questionnaire) and was instructed to use that variety’s CLT comprehension item lists. The experimenter would start by saying the first word on the list and if

83 The BiLI-TAS research team at that stage consisted of Ute Bohnacker (Professor of Linguistics and PI of the project), Linnéa Öberg (PhD student of Linguistics), Anette Månsson (Lecturer in Semitic languages), Karin Koltay (SLP and research assistant) and the author.
the child showed signs of not comprehending the word, the experimenter would ask the question again using the second synonym on the list. Rarely, if the child still showed signs of not understanding, the experimenter would provide a third synonym. It was decided that a child could be given up to three different word options for each item in the Arabic comprehension part, if needed. If the child did provide an answer, regardless of whether it was correct or not, the experimenter would not offer another lexical item, but proceed to the next test item.

Using the Lebanese Arabic version of the CLT in the cross-sectional study in Lebanon was much more straightforward; no further adaptations were needed nor were exceptions to the standard testing procedure necessary. With the Lebanese bilingual children, however, the experimenter constantly encouraged the children to provide the answer in Arabic, rather than in a second language (English/French).

In all studies, the children’s answers were noted on the respective answer sheets for each subtask at the time of testing. The experimenter held a neutral tone and provided minimal feedback (such as aha or mmm) regardless of the answers’ correctness throughout the testing session. If the child did not provide any answer, the experimenter would write no answer or draw a dash (–) in the answer sheet. All answers were noted down regardless of accuracy or relevance. The entire procedure was video and audio recorded to allow for reliable transcription and checks afterwards. For a description of the scoring procedure of the CLT, see Section 4.5.1.

4.2.2 Multilingual Assessment Instrument for Narratives (MAIN)
The Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015; MAIN-Revised; Gagarina et al., 2019) is a narrative elicitation tool that consists of two pairs of picture sequences (i.e. 4 picture sequences in total). Every sequence consists of six pictures depicting a story containing three episodes each with a goal-attempt-outcome sequence. MAIN was originally designed within the COST Action IS0804 (2009–2013) by researchers in Working Group 2 (Narrative and Discourse). The tool was developed to evaluate the narrative production and comprehension abilities of monolingual and bilingual children primarily from ages 3 to 10 across

---

84 With gained experience in testing, the experimenter started to provide two options of synonyms simultaneously depending on which synonyms were most frequent in that variety. (Wen elmunhar aw elanf? ‘Where is the nose or the nose?’). Note that this method does not make the testing procedure easier in the Arabic session than in Swedish, since the child might not have been exposed to the first colloquial form provided. Note also that the child is not given an explanation of the word, but merely another lexical label for the same item. Further colloquial synonyms were given in all cases and for all children when the child showed signs of not comprehending (as described above).
languages. MAIN includes a protocol for assessing production of macrostructural elements (story structure components), and narrative (structural) complexity as well as counting internal state terms (ISTs).

Each MAIN story is presented to the child as a fold-out picture sequence. The pictures are printed in colour on white A4 paper that is cut in half to form a long strip with six pictures. Each of the six pictures is 9 by 9 cm. The content of the MAIN pictures is controlled for aspects concerning macrostructure, characters and their actions and feelings as well as cross-cultural appropriateness and robustness (Gagarina et al., 2015, p.255). The four MAIN stories are parallel in terms of length and story grammar components; however, the number of characters differs from one pair to the other.85

The first pair, the Cat and Dog stories (MAIN1) have identical plotlines and number of characters (3). However, these characters, as well as the objects and general setting, differ between the two stories. The plotline goes as follows: There is a playful cat/dog who wants to catch a butterfly/mouse, it jumps in an attempt to catch its target, but lands unsuccessfully in a bush/on a tree trunk (episode 1). Meanwhile, a cheerful boy comes holding a bucket of fish/bag of sausages and a ball/balloon. The boy, startled when seeing the animals in action, loses his ball/balloon in a nearby lake/tree. He tries to retrieve his ball/balloon and succeeds eventually in his attempt (episode 2). At the same time, the cat/dog sees the boy’s bucket of fish/bag of sausages and decides to eat some. The cat/dog jumps towards the bucket/bag and eats the treat (episode 3).

The second pair, the Baby Birds and Baby Goat stories (MAIN 2) have the same number of characters (5), however, they differ slightly in the plotline with regard to their first episode. In the first episode of Baby Birds, baby birds in a nest are hungry; their mother sees this and flies away to get them food, and shortly returns to feed them. In the first episode of Baby Goats, one of the baby goats is drowning; the mother sees this and goes into the lake in an attempt to save it. The mother succeeds in rescuing her child, whilst the other baby goat is standing aside eating grass (episode 1). The remaining two episodes are quite similar in the two stories. A naughty cat/fox climbs a tree/jumps forward in an attempt to catch a baby bird/goat, and is about to succeed in doing so (episode 2). A brave dog/bird passing by sees the incident and decides to rescue the baby bird/goat. He bites the cat’s/fox’s tail and successfully chases the cat/fox away (episode 3).

All stories are followed by 10 comprehension questions that mainly target the child’s understanding of the characters’ goals and internal states. Table 4.2 summarizes the basic similarities and differences between the two story pairs: MAIN1 (Cat, Dog) and MAIN2 (Baby Birds, Baby Goats). Because the children in the present study were always tested on Cat/Dog first, and second on

85 The MAIN story picture sequences and materials can be downloaded from https://main.leibniz-zas.de.
Baby Birds/Baby Goats, the stories Cat/Dog are referred to as MAIN1, and Baby Birds/Baby Goats as MAIN2.

Table 4.2. Characteristics and components of the four different MAIN stories.

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cat</td>
<td>Dog</td>
<td>Baby Birds</td>
<td>Baby Goats</td>
</tr>
<tr>
<td>Episodes</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Characters</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Main Protagonist(s) in each episode</td>
<td>Ep.1: Cat &amp; Butterfly</td>
<td>Ep.1: Dog &amp; Mouse</td>
<td>Ep.1: Mother bird &amp; baby birds</td>
<td>Ep.1: Mother goat &amp; baby goat 1</td>
</tr>
<tr>
<td></td>
<td>Ep.3: Cat</td>
<td>Ep.3: Dog</td>
<td>Ep.3: Dog, cat &amp; baby birds</td>
<td>Ep.3: Bird, fox &amp; baby goat 2</td>
</tr>
<tr>
<td>Comprehension questions</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>


MAIN is available in more than 60 languages. The language versions of MAIN used in the present study are Swedish and five Arabic versions. The Swedish version was developed by Ute Bohnacker, one of the core authors of MAIN (2012) and MAIN–Revised (2019). Rima Haddad (the author) developed five Arabic versions of MAIN, namely the Lebanese, Syrian, Palestinian and Iraqi (Baghdad and Mosul) Arabic varieties.

The development of the different Arabic versions was carried out in consultation with several native-speaking informants as well as Semitic dialectology experts and linguists. The author (a native speaker of Lebanese Arabic) started by developing the Lebanese version of the comprehension questions (as well as the instructions and prompts) by translating them from English and MSA into Lebanese Arabic and improving them further in consulting with native Lebanese Arabic speakers residing in greater Beirut and Mount Lebanon region.

For the Syrian, Palestinian and Iraqi versions, informants speaking the respective Arabic varieties were consulted. The informants were recruited intentionally to come from different regions and to speak different varieties. The native speakers were asked to study the MAIN picture sequences and then translate the comprehension questions from the English, Swedish and/or MSA versions into their respective Arabic variety. They had also the possibility to

86 For information on MAIN and the different language versions, visit https://main.leibniz-zas.de/en/.
87 The Arabic MSA version was originally translated by Hadil Karawani (2012).
88 The native speakers and the informants who were consulted for the adaptation of the various CLT Arabic varieties also shared their knowledge in the construction of the MAIN comprehension questions. The informants speaking the Syrian varieties spoke those of greater Damascus, Homs province and Jazira province. The informants speaking Palestinian spoke those varieties of greater Jerusalem, Gaza strip and north western Palestine. The Iraqi varieties were those spoken in greater Baghdad, Mosul region and Najaf region. Some of the informants were met in person while others were contacted through e-mail or through voice messages.
look at the Lebanese translation as an example. The Arabic versions were piloted (see Section 4.3.1) and a few necessary amendments were made accordingly. For a fully detailed description of the development of the different Arabic MAIN versions, see Bohnacker and Haddad (2020).

MAIN can be used to elicit narratives in three different modes: *telling* (story generation), *retelling*, or *model story*. In the present study, *telling* mode was used to elicit all four MAIN stories. There were several reasons for this. First, the *telling* mode avoids memory effects that arise from retelling (Boudreau, 2007; Dodwell & Bavin, 2008). Second, employing only the *telling* mode (and not *model story*) enabled us to use all four stories (Cat, Dog, Baby Birds and Baby Goats) to elicit as much narrative possible from the same child in both of her/his languages. Finally, the presence of the various Arabic varieties spoken by the children in Sweden would greatly complicate the use of the *retelling* mode due to lexical, morphological and syntactic (amongst other) differences between the Arabic varieties.

In the present study, standard procedures for testing were used (Gagarina et al., 2019). One language was tested at a time and every child was pre-assigned their MAIN narratives according to a counterbalancing system (see Section 4.3.5 on counterbalancing). Three envelopes containing the same story were placed on a table in front of the child. The child was asked to choose one envelope believing that each one contained a different story. The purpose behind this setup was to make the child think that the experimenter was not aware which story was chosen, to control for effects of shared knowledge and joint attention. The child then took out the pictures from the envelope and looked at all the pictures at her/his own pace before telling the story. When the children were tested in Arabic, they were given picture sequences that started from right to left (to mimic the reading direction of Arabic). If a child started looking at the pictures from the wrong direction (from left to right), the experimenter would remind her/him that the story is in Arabic and that they should therefore direct their attention first to the right. The picture sequence faced the child, and not the experimenter throughout this whole process, giving the impression that the experimenter was still not aware which story the child had chosen. When the child had finished looking at the pictures, the experimenter refolded the pictures so that only the first and the second pictures were visible to the child. The experimenter then asked the child to begin telling the story. Prompts and intrusions by the experimenter were kept to a minimum throughout the testing procedure. The use of prompts was unified as much as possible throughout testing sessions despite the child’s age. Younger children needed at times more prompting than older children, however, the types of

---

89 In the *telling* mode, a child looks at all the pictures and then tells the story unfolding two pictures at a time. In the *retelling* mode, a child first hears the story being told by the experimenter and is then asked to retell the same story. The *model story* mode is similar to the retelling mode, however, the child is asked first to listen to one story and is then asked to tell another story.
prompts used throughout the study were held constant in order to avoid experimenter effects in data collection. Minimal standardized prompts (such as *mmm, and then*) were used by the experimenter if the child stopped telling in the middle of the story. After the child had finished narrating the first two pictures (pictures 1-2), the second pair of pictures (pictures 3-4) was unfolded, making all four pictures visible to the child. Finally, the last two pictures were unfolded (pictures 5-6), making the whole picture sequence visible to the child when s/he was finishing the telling of the story. When the child showed signs of having finished (for example, by remaining silent), the experimenter asked the child whether s/he had indeed finished. On confirmation, the experimenter placed the picture sequence on the table making all the pictures visible to both. The experimenter then proceeded by asking the ten standardized comprehension questions for that story. The entire procedure was audio and video recorded.

When tested in Arabic, each child was asked comprehension questions in the Arabic variety that her/his parents had reported that the child spoke. If a child showed signs of not understanding the question, the experimenter would provide a reformulation of the question using a closely related Arabic variety.90 Note that the experimenter would not change the question per se, but rather use different terminology (synonyms) or word order to ensure that the child had been provided with a question formulation that s/he is familiar with and, as a result, had understood the question. The scoring of MAIN is described in Section 4.5.2.

4.3 Procedure

This section commences with a brief description of the pilot study conducted to try out the Arabic adaptations of the testing materials (Section 4.3.1). It then describes the general data collection procedures implemented in the cross-sectional (Section 4.3.2) and the longitudinal (Section 4.3.3) studies in Sweden and the cross-sectional study in Lebanon (Section 4.3.4). Information regarding the counterbalancing system is found in Section 4.3.5.

4.3.1 The pilot study

In March 2017, a pilot study was conducted with five Arabic-Swedish-speaking bilingual children (age range 6–8). The children were recruited in Landskrona and Malmö (two cities in southern Sweden) by the author via personal contacts. The children spoke either Syrian or Palestinian varieties, and

---

90 This is especially relevant to the questions related to feeling of the protagonists (D2, D5 and D8) where the experimenter sometimes used both versions of the word ‘feeling’ *ḥsās* and *šūʿūr* in the same question formulation (or when the child’s sign of not understanding the question).
were tested only in Arabic. There were several reasons behind conducting the pilot study. First, it was essential to test whether the instructions for each of the tasks in the test battery were comprehensible for the children (regardless of the Arabic variety difference between the child and the experimenter). The second aim was to identify whether the children showed signs of misunderstanding the comprehension questions or target items in the receptive tasks in MAIN and CLT. Finally, the third aim was for the author to gain experience in using the test materials and standardizing data collection procedures. None of the pilot data were included in the present study.

4.3.2 The cross-sectional study in Sweden

One hundred children (ages 4;0–7;11) participated in the cross-sectional study in Sweden. For each child, the same general procedure was followed. Data collection was conducted on two separate occasions, one time in Arabic and one in Swedish, every time by a native speaker of that language. Each session was carried out by a different experimenter, thus creating as much of a monolingual environment with the child as possible, with the aim to elicit data in that the targeted language. The language order of the testing sessions was counterbalanced. Around half of the children were first tested in Arabic and the other half were first tested in Swedish (see Section 4.3.5 on counterbalancing).

The Arabic data collection sessions were conducted by the author (a native speaker of Arabic) and three trained native Arabic-speaking research assistants (Zeinab Shareef, Amal Choumar and Pascale Wehbe). The Swedish data collection was conducted by a native Swedish-speaking PhD student in linguistics (Linnéa Öberg) and a native Swedish-speaking trained research assistant (Karin Koltay). Each data collection session took around 30–45 minutes. The experimenter gave the same instructions to all children.91

The time between the two testing sessions for each child ranged from 2 to 46 days (mean = 10.7 days, median = 7 days being the most frequent time between the two sessions). Differences in the time between the two testing sessions were due to illness, holidays or unpredicted events related to the child or the experimenter.

The testing sessions were conducted in a quiet room with only the child and the experimenter at (pre)school, in the home or at a religious or cultural centre. Originally, it was planned that the Swedish testing sessions take place at school (a place where Swedish is dominantly spoken) and the Arabic testing sessions at the child’s home (a place where Arabic is dominantly spoken). The idea was to meet the child in a setting in which the test language is mostly spoken in order to discourage code-switching between the two languages. This

91 Each task had a prewritten instructions (in the respective language) which the experimenter would carefully follow.
plan needed to be amended since several children were recruited via organizations. Other families preferred that both testing sessions took place at home while others preferred both sessions to take place at school or at the organization. The Arabic testing sessions took place at home ($N = 12$), at (pre)school ($N = 54$) or at a congregation/association ($N = 34$). The Swedish testing sessions occurred at home ($N = 14$), at (pre)school ($N = 54$), at a congregation/association ($N = 31$) or on the premises of Uppsala University ($N = 1$).

In most cases, the experimenter would set up the testing material and recording devices before the child entered the room. When this was not possible, the child often helped in setting up the testing materials, and the experimenter took advantage of this time to bond with the child asking some simple questions providing a friendly relaxed environment. When starting every session, the experimenter made sure to form a connection with the child and to familiarize her/him with the testing situation. After that, the experimenter explained to the child the tasks that they were going to do during the session. The experimenter used a visual timeline consisting of small pictograms printed on adhesive laminated paper. Each pictogram showed a task (parrot/alien for NWR, envelopes for MAIN, and big and small booklets for the CLTs). This method of visualizing the testing session ensured that the child was aware of how many tasks were remaining, and prepared for the testing type of the upcoming task. Removing the relevant pictogram after completing each task gave the child a sense of completion and accomplishment.

The order of testing was identical in both language sessions (Figure 4.3). Before starting each testing session, the experimenter and the child would chat with each other to get to know one another (warm-up). Then, the child would start by doing a non-word repetition task (NWR), then tell a MAIN story (MAIN1: Cat or Dog) and answer the story’s comprehension questions. Next, the child would complete four CLT tasks, followed by telling another MAIN story (MAIN2: Baby Birds or Baby Goats) and answer comprehension questions about it. The session concluded with another NWR task.

<table>
<thead>
<tr>
<th>Warm-Up</th>
<th>NWR</th>
<th>MAIN1 (Cat/Dog) + comprehension questions</th>
<th>CLT (4 sub-tasks)</th>
<th>MAIN2 (Baby Birds/Baby Goats) + comprehension questions</th>
<th>NWR</th>
</tr>
</thead>
</table>

Figure 4.3. Overview of the general structure of the data collection. NWR = Non-Word Repetition task, MAIN = Multilingual Assessment Instrument for Narratives, CLT = Cross-linguistic Lexical Task.

Four different NWR tasks were used to investigate the children’s phonological processing and working memory. The different NWR tasks differed in syllable length, complexity, and whether they were language specific or quasi-universal. The aim behind using these different NWR tasks was to explore the children’s performance across different NWR items with different phonological setup and complexity. Results for the NWR tasks are not reported in the present study; see Öberg (2020).

102
The sessions were video and audio recorded. Whenever possible, the child and the experimenter sat across a table. The audio recorded was placed on the table and the microphone was angled towards the child. The video recording device was placed slightly behind the child to capture the child’s pointing, gestures and facial expressions.

This video setup was not possible at all times due to furniture placement in the testing room. For instance, if the testing was conducted sitting on sofas at home, it was difficult to place the camera behind the sofa (and behind the child). In such cases, referring back to the video recordings was limited. The experimenter did, however, always make sure that she did not see the MAIN picture sequences as the child was telling her/his story regardless of the sitting arrangement, to ensure a non-shared visual setting. Both video and audio recording devices were kept on throughout the whole testing session.

Children were either rewarded with a sticker at the end of each subtask or were given several stickers at once at the end of the whole testing session. The children were always praised (irrespective of actual outcome) and thanked for their participation. At the completion of the second session, each child was also handed a certificate of participation.

4.3.3 The longitudinal study in Sweden

The data collection procedure in the longitudinal study was identical to that of the cross-sectional study in Sweden. Each child was tested in both of their languages on two separate occasions. The same elicitation materials (MAIN, CLT and NWR) were used. Each child was tested following the same counterbalancing order as the one they had in the cross-sectional study.

Ten families from the 4-year-old group in the cross-sectional study agreed to participate in the longitudinal study approximately two years later. The time elapsed between the first encounter (for the cross-sectional study) and the second encounter for the follow-up study ranged between 23 and 28 months. The time between the two language testing sessions of the longitudinal study ranged between 3 and 28 days (mean = 9.8 days, median = 7 days being the most frequent time between the two sessions).

Parents were asked to fill in an informed consent form and a questionnaire. The children were interviewed during the Swedish data collection session, and the parents were interviewed during the Arabic data collection session or shortly after on the phone.

---

93 Due to the COVID-19 pandemic, recruiting a larger number of children for the longitudinal study was not possible. In total 19 families were contacted, out of whom nine families declined to participate due to various reasons, including pandemic restrictions related issues, family relocation or travelling abroad.

94 The aim was to meet the children two years (24 months) after their first testing (i.e. when the children were 6 years old). However, due to national restrictions regarding transportation and gatherings during the COVID-19 pandemic, data collection was postponed until it was feasible.
The Arabic data was collected by the author (a native speaker of Arabic) and the Swedish data was collected by a native Swedish-speaker (Linnéa Öberg, PhD student in linguistics) and a near-native Swedish-speaker (Ute Bohnacker, professor of linguistics). The testing sessions for the longitudinal study took place at the children’s home ($N=3$), at school ($N=5$) or at a religious centre ($N=1$). For one child, the testing sessions were conducted outdoors on park benches outside the home.

In addition to having been rewarded with stickers and a certificate upon completion of the testing sessions, the children received two cinema tickets or a gift card for a bookstore as a gratitude for their participation.

### 4.3.4 The cross-sectional study in Lebanon

The data for the Lebanese cross-sectional study was collected in August 2018 and March 2019. Children were only tested in Arabic, and were thus only seen once (in contrast to two times as in the cross-sectional and longitudinal studies in Sweden). Each testing sessions took around 20–30 minutes. All the testing sessions were conducted by the author, a native speaker of (Lebanese) Arabic, who gave the same instructions to all children. In total, 100 (Lebanese) Arabic-speaking bilingual children aged between 4;0–8;1 were tested.

The testing sessions were conducted in a quiet room at (pre)school ($N=53$), home ($N=23$) or at a congregation/association ($N=24$). Setup procedures regarding the audio and video recording were identical to those of the cross-sectional study in Sweden.

The cross-sectional study in Lebanon did not include any NWR tasks, thus the testing order was as shown in Figure 4.4. Prior to testing, the child and the experimenter chatted with each other to break the ice between them (warm-up). After that, the testing session started. First, the child told a MAIN story (MAIN1: Cat or Dog) and answered the story’s comprehension questions, then completed the four CLT subtasks, and finished by telling another MAIN story (MAIN2: Baby Birds or Baby Goats) and answered comprehension questions on it.

![Figure 4.4](image)

**Figure 4.4.** Overview of the general task order. MAIN = Multilingual Assessment Instrument for Narratives, CLT = Cross-linguistic Lexical Task.

---

95 Due to the national restrictions imposed by the pandemic, many schools stopped opening their doors to outside visitors. (The preschools and schools did not close nor did children have to stay at home since there was no ‘national lockdown’ like in many other countries.) All the testing sessions were, however, conducted respecting the local and national recommendations and restrictions that applied at the time of testing.
Children were rewarded with a sticker at the end of each task or were given several stickers at once at the end of the whole testing session. When the child had completed all tasks, s/he was praised, thanked and handed a certificate of participation.

4.3.5 Counterbalancing

For the cross-sectional study in Sweden, the same counterbalancing procedures were performed for all the tasks, and for both languages (which language was tested first), and for all age groups (Table 4.3). For every 8\textsuperscript{th} child tested, the counterbalancing system was repeated. The aim was to test half of the children in Arabic in their first session and the other half in Swedish in their first session. However, due to differently sized age groups, drop-outs and the exclusion of participants, 53 children were tested in Arabic first, and 47 children were tested in Swedish first. The children also did NWR tasks, two in each language testing session. These tasks were also counterbalanced. See Tables A1.1 and A1.2 in Appendix A1.1 for a more detailed overview on the counterbalancing system of all the tasks as divided per language.

**Table 4.3.** Counterbalancing irrespective of which language the first session started with (Arabic or Swedish).

<table>
<thead>
<tr>
<th>Child</th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NWR</td>
<td>MAIN1</td>
</tr>
<tr>
<td>1</td>
<td>NWR1</td>
<td>Cat</td>
</tr>
<tr>
<td>2</td>
<td>NWR2</td>
<td>Cat</td>
</tr>
<tr>
<td>3</td>
<td>NWR1</td>
<td>Dog</td>
</tr>
<tr>
<td>4</td>
<td>NWR2</td>
<td>Dog</td>
</tr>
<tr>
<td>5</td>
<td>NWR1</td>
<td>Cat</td>
</tr>
<tr>
<td>6</td>
<td>NWR2</td>
<td>Cat</td>
</tr>
<tr>
<td>7</td>
<td>NWR1</td>
<td>Dog</td>
</tr>
<tr>
<td>8</td>
<td>NWR2</td>
<td>Dog</td>
</tr>
</tbody>
</table>

*Note.* BB = Baby Birds. BG = Baby Goats, NWR = non-word repetition task.

The CLT contains four subtasks that were also counterbalanced according to the guidelines by Haman et al. (2015, p.220). Table 4.4 shows the counterbalancing system for noun comprehension, verb comprehension, noun production and verb production according to 4 possible orders.

**Table 4.4.** Counterbalancing system for the four subtasks of the Cross-linguistic Lexical Tasks (CLTs).

<table>
<thead>
<tr>
<th>Order</th>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Noun comp</td>
<td>Verb comp</td>
<td>Noun prod</td>
<td>Verb prod</td>
</tr>
<tr>
<td>2</td>
<td>Verb comp</td>
<td>Noun comp</td>
<td>Verb prod</td>
<td>Noun prod</td>
</tr>
<tr>
<td>3</td>
<td>Noun prod</td>
<td>Verb prod</td>
<td>Noun comp</td>
<td>Verb comp</td>
</tr>
<tr>
<td>4</td>
<td>Verb prod</td>
<td>Noun prod</td>
<td>Verb comp</td>
<td>Noun comp</td>
</tr>
</tbody>
</table>

*Note.* comp = comprehension, prod = production.
The same counterbalancing procedures were applied irrespective of which language the child started with in their first session.

To give a concrete example of how the counterbalancing system was applied in the cross-sectional study in Sweden, consider child number 1 in Table 4.3. If this child started her/his first session in Arabic (session 1), then that child’s testing session started with doing a NWR task, followed by telling MAIN1 Cat story (and answering the comprehension questions), then doing the CLTs according to test order 1, then telling MAIN2 Baby Birds story (and answering the comprehension questions) and finally doing a second NWR task, all in Arabic. When that same child was tested the second time (session 2), this time in Swedish, the child started with doing a NWR task, then telling MAIN1 Dog story (and answering the comprehension questions), then doing the CLTs according to test order 1, followed by telling MAIN 2 Baby Goats story (and answering the comprehension questions) and finally doing a second NWR task.

As for the longitudinal study, the children were tested following the same test order they had been tested with in the cross-sectional study. The reason for this was to observe each child’s performance at the two ages on exactly the same tasks. As a result, 4 children were first tested in Arabic and 6 children were first tested in Swedish.

In the Lebanese cross-sectional study, the children were only tested in Arabic for their narrative (using MAIN) and lexical skills (using CLT). The counterbalancing system was repeated for every 4th child, as displayed in Table 4.5.

Table 4.5. Counterbalancing system for the cross-sectional study in Lebanon.

<table>
<thead>
<tr>
<th>Child</th>
<th>MAIN1</th>
<th>CLT</th>
<th>MAIN 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cat</td>
<td>1</td>
<td>Baby Birds</td>
</tr>
<tr>
<td>2</td>
<td>Cat</td>
<td>2</td>
<td>Baby Goats</td>
</tr>
<tr>
<td>3</td>
<td>Dog</td>
<td>3</td>
<td>Baby Birds</td>
</tr>
<tr>
<td>4</td>
<td>Dog</td>
<td>4</td>
<td>Baby Goats</td>
</tr>
</tbody>
</table>

4.4 Transcription

All of the children’s lexical production answers, narratives and answers to the MAIN comprehension questions were transcribed. The CLT production responses (nouns and verbs) were written down by the experimenter on scoring sheets at the time of testing. The children’s Swedish responses were written using standard Swedish orthography. Arabic responses were either written in standard Arabic orthography or in Latin letters and numbers (see description below), whichever method was most convenient for the experimenter. When inserting the lexical production answers into Excel sheets later, Latin transliteration was used throughout and video and audio recordings were used to
check unclear responses. All the unclear responses were evaluated by several native or near-native speakers of the respective language (Arabic or Swedish) on the BiLI-TAS team.

As for the narratives in the cross-sectional study in Sweden, each child told four MAIN stories, two in each language. The children’s narratives were transcribed verbatim from the audio recordings, then checked and expanded using the video recording. Further information regarding non-verbal responses (for example, pointing or gesturing) was added in the video check. Transcriptions were then rechecked several times (see below).

Transcriptions were made following the BiLI-TAS project transcription guidelines (Guidelines for the Transcription of MAIN, Bohnacker, 2018c) which are based on the CHAT transcription format (MacWhinney, 2000). See Appendix A1.3 for examples of MAIN transcripts in Arabic and Swedish.

Conversational turns were used as the basic unit for transcription. A conversational turn of one speaker ends when the other speaker starts speaking. Hence, interruptions or comments by the experimenter also determined a conversational turn.

Trained transcribers made several passes over all transcribed narratives. The transcriber let the transcript rest for a certain amount of time and then listened to the recording again. The Arabic transcriptions were made by the author (a native speaker of Arabic) and two trained native Arabic-speaking research assistants (Pascale Wehbe and Zeinab Shareef). The Swedish transcriptions were made by two experienced transcribers, one native speaker of Swedish (Linnéa Öberg) and one trained near-native Swedish-speaking research assistant (Frauke Jonsson). All the transcriptions were rechecked in their entirety against the video files after around one year of the original transcript. Amendments were made and noted. In some cases, the author also listened to the Swedish narratives of some children who had several instances of Arabic influence while narrating. Comments were added to the Swedish transcripts regarding the meaning of Arabic code-switched words.

As previously mentioned, for the Swedish transcripts, standard orthography was used and for the Arabic transcripts, Latin letters and numbers were used. The author developed a unified transcription system for the Arabic narratives (Haddad, 2019). Since the aim for the transcriptions was to analyse

---

96 In a few cases (N = 5), video recordings were not available due to technical failure. For these, non-verbal information was not recorded.

97 The Arabic transcription guidelines were developed by the author (2018–2019) after discussions on transcription possibilities with Rachel Fiani, Nouhad Abou Melhem, Christel Khoury Aouad Saouk, Mahmut Aghbat, Zeinab Shareef, and Pascale Wehbe in addition to Ute Bohnacker and Anette Månsson. The result was a transcription system that was fast and suited the needs of the project. The orthography for the transcription system was inspired by the commonly used method of communication on internet platforms using Latin letters and numbers. Unification rules on writing certain sounds (for example, vowels) were constructed. Throughout the dissertation, however, Arabic examples are transcribed following Eid et al. (2011).
the children’s narrative macrostructural production and comprehension, emphasis was not put on capturing the phonological accuracy of the utterances, but rather on understanding the child’s lexical word choice. The transcriber thus aimed to unify the orthography across the Arabic varieties but did not change the child’s lexical choice from colloquial Arabic to Modern Standard Arabic (MSA). For instance, the word ‘cat’ differs in several Arabic varieties (for example, bsayne, bisse, qitta, hirra, bazzuni). When one of these words occurred in a child’s narration, the words would be transcribed as it is (the transcriber would not replace one lexical choice with another, i.e. the word bsayne ‘cat’ would not be replaced with qitta (‘cat’ in MSA)). The transcriber would, however, use the same spelling, regardless of the child’s change in vowels (for example, the word bisse ‘cat’ was sometimes also uttered as bissa, but would be transcribed as bisse at all times). A list of pre-specified spellings of the most common nouns and verbs in the MAIN narratives was prepared and followed as the Arabic narratives were transcribed.

As a general transcription rule, words were transcribed orthographically and not phonetically. Words that were uttered with great phonological deviations from the original pronunciation, were transcribed as heard by the transcriber with a potential explanation within squared brackets and colon [: ] following CHAT conventions. For example, the slight phonological deviation (e instead of a) of a 4-year-old child (BiAra4-11) in the Swedish cross-sectional study when telling MAIN1 (Cat story) could with ease be interpreted for what the child originally intended to say: han vill (.) att henne titter [: tittar] ‘he wants (.) that her looks’.

Repetition, rephrasing (retracing) and reformulation were coded with [/], [//] and [///] respectively. Fillers, interrupted or unfinished words, and non-linguistic sounds (such as imitating animal noises) were coded using the ampersand &-symbol. When these symbols are added in the transcribed texts, the Computerized Language Analysis program (CLAN: MacWhinney, 2000) can automatically exclude these items from calculations of word types and word tokens. For example, the hesitation filler ‘ehh’ would not be counted as a unique word type when transcribed with the ‘&’ symbol placed in front of it (&ehh). Instances of code-switching to a language other than that of the testing session were marked using the codes [@s] for Swedish, [@a] for Arabic, and [@e] for the rare occasional English. (This coding enables CLAN to exclude or include the code-switched words in any analysis.) Certain loan words (such as balon, ‘balloon’) were not transcribed as code-switches but considered as part of the Arabic language. Children may not be familiar with the Arabic equivalent for certain words since the loan word is frequently used by adults amongst themselves and to the children.

98 A unified transcription system was used to simplify word counts (word tokens), analysis of the number of different words (NDW) and allow for automated searches for individual words, regardless of the Arabic variety that each child speaks.
Transcriptions were completed by passing through six steps.

- First, transcriptions were made from the audio files in the EXMARaLDA (Schmidt & Wörner, 2014) transcription program. Each audio file was cut into several small conversational turns which made it easier for the transcriber to repeat small audio recordings several times when needed. Different conversational tiers were used for the child and for the experimenter, in addition to one comment tier, following CHAT conventions.99

- Second, the transcriptions were checked against the video recordings, in order to add comments regarding non-verbal information such as nodding and other kinds of gestures to indicate a yes/no/I don’t know, pointing, gesturing, unfolding of the pictures and the use of special sounds (for example, the change in a child’s voice or use of onomatopoeic sounds). Comments regarding unusual circumstances or occurrences were also added (for instance, if someone entered the room or whether abrupt interruptions occurred during the narration). Additional information was noted at the beginning of the transcript concerning general observations regarding the child’s behaviour (if the child seemed unfocused or tired), whether the child had articulation difficulties, information whether there was another person in the room, or whether the recorded files were hard to listen to (usually, due to background noise).

- Third, the transcriptions were exported from EXMARaLDA to CHAT-format so that they could be analysed in CLAN.

- Fourth, transcription file headers were added to the CHAT-format texts. The headers included information regarding the child’s individual code, age and gender of the child, name of the experimenter, name of the transcriber, which narrative was told (MAIN1 Cat/Dog, MAIN2 Baby Birds/Baby Goats) and the date of the test session.

- Fifth, the transcriptions were rechecked in both languages (by the author for the Arabic transcriptions, by Linnéa Öberg for the Swedish ones) against the video recordings. The Arabic orthography was unified as much as possible. Unclear cases were discussed and solved, and comments regarding code-switching were added.

- Sixth, a final check for reliability was made for Arabic and Swedish MAIN transcriptions as detailed below. The reliability check was more extensive for Arabic because of the various varieties that were spoken by the children in the study.

---

99 At rare times, another adult, such as a parent, would be sitting in the room. If the parent interfered with the testing session, this person would have a separate tier and their speech would be transcribed, followed by a comment.
In Arabic, the author rechecked all the Arabic MAIN1 transcriptions and the 4-year-olds’ MAIN2 transcriptions. The transcripts were checked against audio and video recordings. Iraqi speaking (SLP, and trained research assistant) Zeinab Shareef was consulted when the audio recordings of Iraqi speaking children were incomprehensible for the transcriber or the checker. Pascale Wehbe rechecked all the remaining MAIN2 transcriptions (for the 5-year-olds, 6-year-olds, and 7-year-olds) as well as the transcriptions from the longitudinal study. The transcripts were checked against the audio recordings. In general, the changes were few and included mostly minor changes and spelling changes. Seventeen utterances initially transcribed as ‘xx’ (unintelligible sound) were changed and corrected. In Swedish, Ute Bohnacker checked the MAIN1 and MAIN2 transcriptions of 13 randomly selected children (3 children from each age group, and 4 children from the 6-year-old group since this age group had relatively more children than the others). The transcripts were checked against audio and video. Word counts (for the child’s utterances only) were performed in CLAN. The percentage of agreement (inter-transcriber reliability) was 98.3% for MAIN1 and 97.9% for MAIN2.

4.5 Scoring

The BiLI-TAS research team created two scoring guidelines, one for the scoring of CLT (Guidelines for Scoring CLT, Bohnacker 2018a) and one for MAIN (Guidelines for Scoring Macrostructure in MAIN, Bohnacker 2018b). The aim was to homogenize scoring across languages, children and scorers as well as to provide rationales for the scoring decisions so that scoring can be replicable. The scoring guidelines were written and continuously revised (2015/2016–2018) as the BiLI-TAS dataset in Swedish, English, German, Turkish (and most recently Arabic) was growing. After extensive discussions and consensus rounds, the team formed principles and guidelines for the scoring of the CLT and the MAIN.

As a general rule concerning the scoring of all tasks, if a child’s response was uttered or self-corrected as a result of the interference of another person (for example, an adult or a parent present in the room), then that answer would not be scored (this occurred very rarely). The aim is to test the child’s knowledge rather than a refined corrected version of it. If the child self-corrected without any assistance, then the new answer would be scored, hence, the initial response would be ignored regardless whether it was correct or not.

100 The BiLI-TAS team at that point consisted of Ute Bohnacker (Professor of Linguistics and PI of the project), Sibylle Dillström (SLP, PhD in German; researcher), Josefin Lindgren (PhD student of Linguistics), and Buket Öztekin (SLP, PhD student of Linguistics), in addition to Karin Koltay (SLP, research assistant) who was present in the forming of the MAIN Guidelines.
4.5.1 Scoring of CLT

The scoring of the CLT was based on the BiLI-TAS CLT scoring guidelines (Guidelines for Scoring CLT, Bohnacker, 2018a). Scoring of the Arabic CLT was done by the author, a native speaker of Lebanese Arabic. The author additionally consulted Ute Bohnacker, professor of linguistics, and Anette Månsson, lecturer in Semitic languages, when scoring the Arabic production responses. Scoring of the Swedish CLT was done by Linnéa Öberg, a native Swedish-speaking PhD student of linguistics, and unclear cases were discussed by the BiLI-TAS team.

For each correct answer, one point was awarded. The maximum score for CLT comprehension is 60 points (30 points on noun comprehension and 30 points on verb comprehension). Likewise, the maximum score for CLT production is 60 points (30 points on noun production and 30 points on verb production). All children in the 3 studies performed all four CLT sub-tasks. In total, 38,400 responses were scored, as explained in Table 4.6.

**Table 4.6.** Number of scored CLT responses across the studies.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study in Sweden</th>
<th>Longitudinal study</th>
<th>Cross-sectional study in Lebanon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of test items</td>
<td>120 test items x 2 languages</td>
<td>120 test items x 2 languages</td>
<td>120 test items x 1 language</td>
</tr>
<tr>
<td>Number of children</td>
<td>100</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Number of scored CLT responses</td>
<td>24,000</td>
<td>2,400</td>
<td>12,000</td>
</tr>
</tbody>
</table>

*Note.* The 120 test items consist of 60 comprehension and 60 production items.

4.5.1.1 CLT comprehension

In the CLT comprehension task, when the child pointed to (or gave the number of) the correct picture (picture identification), one point was awarded. At the time of testing, the experimenter would write down on the scoring sheet the number of the picture that the child had designated. If the child pointed to one picture and then changed her/his mind, the second answer would be scored. If the child pointed to two pictures, then s/he was asked to choose only one of the pictures, and the latter chosen picture was scored.

4.5.1.2 CLT production

All uttered responses were written down by the experimenter on the scoring sheet at the time of testing. All responses were then transcribed into an Excel sheet. If the experimenter had jotted down a remark or had written unintelligibly on the scoring sheet, the scorer would refer back to the audio/video

---

101 When given the instruction for the task, the children were asked to reply to the prompt questions (*what is this?* or *what is s/he doing? / what is happening here?*). Common wrong replies included providing descriptions, explanations, examples or making gestures. The children’s utterances were all noted down by the experimenter.
recordings for consultation. As a next step, the child’s response would be scored (0 or 1). See below.

The scoring of the Arabic CLT production task involved several steps due to the presence of various Arabic varieties in the data. A large number of lexical labels emerged in addition to the Lebanese Arabic target answers. Thus, in order to score the Arabic production part, the correctness of certain ‘new’ lexical labels (i.e. labels in additional Arabic varieties) needed to be validated. The author consulted Ute Bohnacker and Anette Månsson, Semitic scholars, Arabic native speakers as well as parents of children participating in the present study. This labour intensive procedure was implemented to ensure that no correct response in any Arabic variety (that differed from the target label in Lebanese Arabic) was mistakenly scored as incorrect.

As for the scoring of the Swedish CLT production, only few new tricky cases emerged as most were already covered in the previously established guidelines. Tricky cases were solved through discussions with the BiLI-TAS team until agreement was reached.

Apart from producing the correct target word, certain additional responses were also scored correct:

1. Errors in grammatical gender (e.g. *ett tiger* instead of *en tiger* ‘a tiger’ in Swedish)
2. Marking of tense, finiteness, person or plural; as long as the child used the target lemma it was scored as correct (e.g. *dansar* ‘dancing’ and *dansa* ‘to dance’ in Swedish; *bašle* ‘an onion’ and *bašal* ‘onions’ in Arabic)
3. Pronunciation of target words that deviates from adult-like pronunciation, mainly due to age-appropriate phonological processes or colloquial variation (e.g. *ruschingkana* instead of *rutschkana* ‘a slide’ in Swedish; *samsiyya* instead of *šamsiyya* ‘an umbrella’ in Arabic)
4. A more specific label than the target word corresponding to the picture (e.g. *gummistövlar* ‘rubber boots’ instead of *stövlar* ‘boots’ in Swedish)
5. Alternative adult-like interpretations of the picture that are colloquial or sociolectally acceptable (e.g. *warda* ‘rose’ instead of *zahra* ‘flower’ in Arabic)
6. Adult-like near-synonyms of the target word corresponding to the picture (e.g. *måla* ‘to paint’ instead of *rita* ‘to draw’ in Swedish).

Responses that were scored as incorrect included:

1. Wrong part of speech (e.g. *vatten* ‘water’ (noun) instead of *vattna* ‘to water’ (verb) in Swedish on verb production, *sura* ‘a picture’ (noun) instead of *šawwar* ‘took a picture’ (verb) in Arabic on verb production)
2. Code-switching, producing words in another language instead of the target language (e.g. *måla* in Swedish instead of *lawwan* ‘coloured’ in Arabic)
3. Phonologically too deviant responses that are barely recognizable as the target word (e.g. *kringer* instead of *springer* ‘runs’ in Swedish, *samsa* instead of *šamsiyye* ‘umbrella’ in Arabic)

4. Phonologically deviant responses (mainly due to metathesis) that result in a different lexical entry than the target word (e.g. *fixa* [ˈfiksə] ‘to fix’ instead of *fiska* ‘to fish’ in Swedish; *sahab* ‘pulled’ instead of *sabah* ‘swam’ in Arabic)

5. A more general word than the target item (e.g. *jobba* ‘work’ instead of *bygga* ‘build’ in Swedish, *nayim* ‘sleeping’ instead of *yitšammas* ‘tanning (on a sunbed)’ in Arabic) or a less specific word than the target word (e.g. *stāda* ‘to clean’ instead of *sopa* ‘to sweep’ in Swedish) or general multipurpose verbs (e.g. *yaʾmil ḥufra* ‘make a hole’ instead of *yahfar* ‘dug’ in Arabic)

6. Paraphrases, circumlocutions or explanations instead of saying the target word (e.g. *lägger papper i lådan* ‘put paper in the box’ instead of *posta* ‘to post’ in Swedish; *ḥatt may bi elkibbeye* ‘put water in the glass’ instead of *sakab* ‘poured’ in Arabic).

Further detailed information regarding scoring of the CLT can be found in the BiLI-TAS scoring guidelines (Bohnacker, 2018a).

### 4.5.2 Scoring of MAIN

The children in the cross-sectional and longitudinal studies in Sweden performed MAIN in both Arabic and Swedish. This yields a potential total of 7480 scorable macrostructure components in the narrative production part and 4400 scorable responses in the narrative comprehension part, as shown in Table 4.7.

<table>
<thead>
<tr>
<th>Table 4.7. Number of scored MAIN responses across the studies in Sweden.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional study</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>MAIN Production</td>
</tr>
<tr>
<td>MAIN Comprehension</td>
</tr>
<tr>
<td>All potential MAIN responses</td>
</tr>
<tr>
<td>(100 children) x 17 Macro (x 4 stories)</td>
</tr>
<tr>
<td>6800</td>
</tr>
<tr>
<td>(10 children) x 17 Macro (x 4 stories)</td>
</tr>
<tr>
<td>680</td>
</tr>
</tbody>
</table>

*Note. Each child had 4 chances (2 in each language) for doing MAIN production and MAIN comprehension. Macro = macrostructural components.*

In each language, a handful of children were not able to perform the narrative task and/or answer the comprehension questions. See Chapter 7 for a detailed overview on the number of children who did not answer the comprehension questions (i.e. have missing narrative comprehension scores) and for the number of children who did not tell the narrative (i.e. have missing narrative production scores).
The scoring of MAIN followed the BiLI-TAS MAIN macrostructure scoring guidelines\textsuperscript{102} (Bohnacker, 2018b), which are a detailed predecessor of the MAIN-Revised scoring protocol (Gagarina et al., 2019).

The author, a native speaker of Arabic and a near-native speaker of Swedish, scored all MAIN macrostructure production and narrative comprehension responses in both Arabic and Swedish. All Swedish MAIN comprehension responses were independently scored by Ute Bohnacker as well. Tricky cases and all instances of interrater disagreement were highlighted and discussed further until agreement was reached. For each score, the rationale for the scoring and date was documented.

Scoring and checking against the guidelines was done in several rounds and at different time points. The entire Swedish production scoring were checked again against the guidelines by Linnéa Öberg. Disagreements were noted and discussed with Ute Bohnacker and the author. Interrater agreement was 97.15%. The entire Arabic production scoring were jointly rechecked again against the guidelines by Ute Bohnacker and the author. After intensive discussions and consultation of audio and video recordings, rationales for the scoring and the date of the scoring decision were documented.

For the scoring of unresolved tricky cases in both languages, additional rounds of discussion between the BiLI-TAS team members took place until consensus was reached. Despite the thousands of examples of responses to the comprehension questions available in several languages (English, German, Turkish, Swedish) gathered by the BiLI-TAS team, new responses and tricky cases emerged from the present data that needed to be discussed by the research team. The BiLI-TAS team at that time consisted of Ute Bohnacker, Linnéa Öberg, Frauke Johnsson, and the author. The rationale for the score of such items was documented.

4.5.2.1 MAIN narrative comprehension

MAIN comprises four stories (Cat, Dog, Baby Birds, Baby Goats); each story is accompanied by 10 comprehension questions targeting the understanding of goals, internal states, and the general plotline of the story. Directly after the child had told the story, the experimenter posed the comprehension questions while pointing to certain characters or pictures relevant to the respective questions.

For every correct answer, one point was awarded. A child could thus score up to 10 points on the comprehension task for each MAIN story. As an example, Table 4.8 displays the comprehension questions for the Cat story (MAIN1). The MAIN narrative comprehension questions for all four stories

\textsuperscript{102} The BiLI-TAS MAIN scoring guidelines contain examples for macrostructure production and narrative comprehension from 1000 transcribed texts and around 9040 comprehension question responses gathered during the MAIN pilot studies and during the BiLI-TAS project, further investigated in Bohnacker (2016), Lindgren (2018), and Öztekin (2019).
can be found in Appendix A1.4 and are available in MAIN-Revised (Gagarina et al., 2019).

Questions D1, D4 and D7 examine the child’s understanding of the goals of the main characters in each of the three episodes. Questions D2 and D5 are related to the internal states of the characters of Episode 1 and Episode 2 respectively. Question D8 is a theory of mind question that examines the child’s ability to think of a hypothetical situation that does not occur in the actual story but that requires the understanding of the plot. Questions D3, D6 and D9 are follow-up questions to D2, D5 and D8 respectively and are only scored if the preceding questions are correct following standard procedure (Gagarina et al., 2019). 103 Question D10 is an overall plotline question that contains a ‘yes/no question’ and a follow-up ‘why question’. One point is awarded if both questions in D10 are answered correctly.

<table>
<thead>
<tr>
<th>Table 4.8. The ten comprehension questions for Cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
</tr>
<tr>
<td>D1. Episode 1 Goal</td>
</tr>
<tr>
<td>D2. Episode 1 IST</td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
</tr>
<tr>
<td>D5. Episode 2 IST</td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
</tr>
<tr>
<td>D8. Episode 3 IST Theory of Mind</td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale Theory of Mind</td>
</tr>
<tr>
<td>D10. Overall plotline</td>
</tr>
</tbody>
</table>

Note. IST = Internal State Term.

The scoring of the MAIN comprehension questions was based on whether the answers were correct and whether they were formulated in a comprehensible manner. The answer does not have to be grammatical or lexically correct, however it should be understandable to the listener/scorer. Hence, the child’s linguistic production competence and knowledge of the respective language played a role. Answers that showed that the child had understood the question

103 Questions D3, D6 and D9 are why-questions that probe the reason behind the character’s internal state. *Why does the x feel...?* These questions are only posed to the child if the preceding questions were answered correctly. If the experimenter asked the follow-up question despite the child had given a wrong answer to the preceding question, then the follow-up question would not be scored correct, regardless of its content.
yet were purely expressed non-verbally (for instance, through miming, gesturing or screaming) were not awarded points. Since the stimulus pictures are available for both child and experimenter, a combination of pointing and a verbal response is acceptable. For example, for question D10 (‘Will the boy be friends with the cat/dog?’), merely nodding or shaking the head as a reply for the ‘yes/no question’ is acceptable if the child provides a correct rationale in the answer to the follow-up question (‘Why?’), e.g. ‘because the cat/dog ate the boy’s fish/hotdog’.

Answers that were provided in another language than the target language were not awarded points. If the child answered in the target language but also included code-switching of certain words, scoring of such answers would be done by ignoring the codeswitched words. If the answer was still comprehensible in the target language, then a point would be awarded.

Lenient scoring was applied for unclear referencing. The child was not ‘punished’ for not providing an adult-like referent encoding, as long as the answer was comprehensible in the situational context, with the pictures being visible to both child and experimenter.

### 4.5.2.2 MAIN narrative production

Each one of the four MAIN stories consists of three episodes. A child is awarded one point for the successful mentioning of each of the five macrostructural components that make up these episodes, namely the internal state as initiating event, goal, attempt, outcome and internal state as a reaction. Additionally, a child is awarded two points if the setting of the story is mentioned (one point for the mentioning of time and one point for place). Thus, the maximum score obtained in macrostructure production is 17 points per narrative. Table 4.9 illustrates setting and macrostructural components in Episode 1 of Baby Birds (MAIN 2).

Table 4.9. Examples of setting and macrostructural components of the first episode of Baby Birds.

<table>
<thead>
<tr>
<th>Macrostructural component</th>
<th>Baby Birds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>Long ago / Once upon a time … (time) … in a bird’s nest / up in a tree (place)</td>
</tr>
<tr>
<td><strong>Episode 1</strong></td>
<td></td>
</tr>
<tr>
<td>IST as Initiating event</td>
<td>Baby birds were hungry</td>
</tr>
<tr>
<td>Goal</td>
<td>Mother bird wanted to feed the babies</td>
</tr>
<tr>
<td>Attempt</td>
<td>Mother bird flew away to look for food</td>
</tr>
<tr>
<td>Outcome</td>
<td>Mother bird brought back food / fed the babies</td>
</tr>
<tr>
<td>IST as Reaction</td>
<td>Baby birds were happy / not hungry any more</td>
</tr>
</tbody>
</table>

*Note. IST = Internal State Term.*

Each of the child’s narratives was scored according to two different methods. The first method consisted of scoring based on the maximum points achieved in producing macrostructural components, max = 17 points in a narrative. The
second method of scoring captures macrostructural complexity (story complexity); namely, whether the child produced sequences of macrostructural components within an episode. Following the MAIN protocol, these component sequences were classified into four categories: attempt-outcome (AO), goal-attempt/goal-outcome (GA/GO), complete episode consisting of goal-attempt-outcome (GAO) and ‘no sequence’.

Similarly to the scoring of the macrostructure comprehension part, in order to receive a point, the child had to answer verbally in a (relatively) clear manner. Non-verbal responses (such as pointing or gesturing) were not taken into consideration when scoring. Recall that in the narration part, only the child was able to see the picture sequences, and not the experimenter. Also, narrating using words in the non-target language (for example code-switching to Swedish while narrating in the Arabic session) would not be taken into consideration when scoring. An utterance would only be awarded a point if it contained a macrostructural component that was still comprehensible after removing the code-switched words in the narration. Non-adult-like use of reference was dealt with leniently, so that it would not negatively affect macrostructural scoring. As long as an utterance encoding a macrostructural component was clear enough who it was referring to, a point was scored despite the non-adult like use of reference. Below is an example of an acceptable case on lenient scoring regarding non-adult like referencing:104

(2) en varg vill äta upp den där (‘a wolf wants to eat that one up’)  
>> och vargen tar den där (‘and the wolf takes that one’)  
>> och den där blir rädd för vargen (‘and that one gets scared of the wolf’)

In example (2), the child introduces the baby goat with den där (‘that one’) without lexically specifying what it refers to. The child repeatedly refers to the baby goat (and only the baby goat) with den där throughout the narration.

Odd lexical choices, vague semantics, ungrammatical utterances or utterances with omissions were scored based on whether the utterance was still understood by the scorer(s) and whether it corresponded to a story component. The child had to describe an action or an emotion that could be linked to a specific character. This means that verb choice had to be relatively specific (as opposed to the vague use of general verbs such as do, go or come without any further specification). Lexical and grammatical errors did not affect macrostructure scoring as long as the utterance was comprehensible either by itself and/or in context with the surrounding utterances. Cases when the choice of lexicon was wrong/vague or when grammatical choices led to misleading meaning, were not scored correct through. For example, in Episode 1 of Baby Birds (MAIN2) recall Table 4.9, if a child said ‘Mother bird comes,’ this would have scored zero points for Outcome since the verb come is too vague

104 The example is from Baby Goats in Bohnacker (2018b).
and does not express the specific aspect of the mother bird bringing back food to feed her babies.

4.6 Ethics considerations
All the parents of the participating children were informed both via a leaflet and the informed consent form in Arabic and Swedish that their child would be video and audio recorded for research purposes, and that confidentiality and anonymity of the child and the parents would be preserved. The leaflet included information describing the nature and objectives of the study, the inclusion criteria for participation and the ability to withdraw from participation at any time without the need to justify their decision. The parents returned the informed consent form signed prior to the first testing session. In a couple of cases, the parents only gave oral consent first, and the signed consent form was returned at the time of testing or soon thereafter.

No separate recording permission was asked from the children since the children were under age and since their parents had already granted permission. However, the children were made well aware of the presence of the video and the audio recorder during the testing session. The families were also offered a copy of their child’s video and audio recordings, if so desired. The parents of the children in the longitudinal study were especially interested in obtaining the video recordings. Video recordings of both language sessions at both testing times (at age 4 and age 6) were highly appreciated.

Video recordings, in addition to the audio recordings, were needed in the present study in order to have information on where the child points. For example, several of the younger children (especially those with low a level of language proficiency or those who were timid during the testing session) preferred to point to a picture in the vocabulary comprehension tasks rather than name the number of the picture, or to gesticulate with their hands in the production tasks when they did not know the target word. Similarly, during the narrative tasks, video recordings are essential in order to identify whether the child points at a character or an object while narrating or when answering the comprehension questions. Video recordings also help to show what the child is saying, especially when audibility is hampered in the audio recording (usually due to child mumbling or due to outside noise). Hence, looking at the movement of the child’s lips helps at times to disambiguate unintelligible speech. Video recordings are also important for analysing whether there was

---
105 The audio recorder was placed on the table with the microphone directed towards the child. The camera was often placed beside or slightly behind the child. No child expressed any objection to being video or audio recorded. Children often helped in setting up the camera and waved ‘hello’ to it.
any unusual nonverbal behaviour by the child, the experimenter or the sur-
roundings during the session. For example, the child might appear to be bored,
tired or distracted by nearby pictures or stickers, or the experimenter might
deviate from the protocols of the tasks. All of the above mentioned are hard
to determine or analyse with merely an audio recording.

Prior to testing the children in the Arabic session, the experimenter would
find out which Arabic variety the child spoke and accordingly use the most
appropriate variety. Pre-prepared prompts, questions and elicitation phrases
were used to match the child’s Arabic variety. The experimenter tried as much
as possible to speak the Arabic variety spoken by the child, and always re-
minded the child to signal if s/he did not understand. This was done to mini-
mize the possibility that any child would be unduly disadvantaged because of
their variety of Arabic.

The general atmosphere during the testing session was relaxed and inviting
for a conversation. Children had fun and were awarded stickers by the end of
each subtask or an abundance of stickers at the end of the session. The children
were praised, regardless of the result, and were given a colourful certificate of
participation (designed by the author) as a gratitude for their efforts.

As previously mentioned (Section 4.1), all children were given individual
codes to preserve their anonymity in research presentations, published mate-
rials, including the present dissertation, and during internal project discussion
meetings.

Transcripts, scoring sheets and coded data have been anonymized so that
the child cannot be identified, and are kept in locked rooms and on servers not
accessible to unauthorized individuals. Information about the children in pa-
paper format (informed consent forms and questionnaires) were organized in
folders that were stored in locked rooms not accessible to unauthorized indi-
viduals. Information preserved in digital form (video and audio recordings,
transcripts and data scores) were stored and backed up on external hard drives
also in locked rooms not accessible to unauthorized individuals. Digital files
containing personal information about the participants were preserved on
servers with access granted only to three individuals working on the study, the
project manager, the author, and another PhD student.

The consent forms and filled-in questionnaires of families who decided to
withdraw from the study were destroyed and all electronic information about
them was deleted.

At the outset of the BiLI-TAS project (2014), the advisory board of the
project discussed the type of data to be collected and the Uppsala University
ethics recommendations and Swedish legislation in place at the time. It was
determined that the data did not involve what according to SFS 2003:460 is
called ‘sensitive information’, and therefore no ethics approval was required
from the regional ethics approval board. The Faculty of Languages (Uppsala
University) did not at the time have an institutional review board either. Nev-
evertheless, the members of the BiLI-TAS project were well aware there are
ethical issues with this type of research project. That is why the above men-
tioned procedures were put in place.

The project manager and principal investigator of the BiLI-TAS project
also strove to ensure that all researchers and assistants involved in the present
study, as well as the larger BiLI-TAS project that this study is part of, followed
good ethical standards throughout. The experimenters (the author, another
PhD student of linguistics, a professor of linguistics and the research assis-
tants) were well trained in advance in how to use the materials and how to
work with children. All of the experimenters had previous experience in work-
ing with children, either from their profession as an SLP, their profession as a
schoolteacher and/or from previous extensive children’s data collection.
5. Participants

In this chapter, the participants are characterised in light of the background information we have concerning their language input and use. First, the participants of the cross-sectional study in Sweden (Section 5.1) are described, followed by the participants of the longitudinal study in Sweden (Section 5.2) and ending with the participants of the cross-sectional study in Lebanon (5.3). As previously mentioned in Chapter 4, information was provided by the parents of the children via an extensive questionnaire. Additional information concerning the participants of the longitudinal study was also gathered via a parental interview and an informal child interview.

5.1 Participants of the cross-sectional study in Sweden

5.1.1 General information

One hundred Arabic-Swedish-speaking children between the ages of 4;0-7;11 participated in the cross-sectional study in Sweden.

Seventeen additional children were originally recruited yet were excluded from the study: five of these children were excluded (after finishing both sessions) since they could not complete all tasks in one or both languages. Additionally, two children were excluded (after completing both sessions) since they spoke Sudanese Arabic varieties. Mutual understanding in the Arabic session was hampered as Sudanese was too distant from the experimenter’s Arabic variety and from the prepared Levant and Iraqi Arabic versions of the data collection materials. Seven children were excluded (after completing their first session) since they were not able to carry out the tasks in the first testing and showed minimal signs of comprehension or participation. One child was excluded (after the first session, in Arabic) since it was revealed that the child was not an active bilingual in Swedish as she attended a French-medium school. Finally, two children were excluded since there was a miscommunication about the child’s age and they had not turned 4 years at the

106 In some cases, the filled-in parental questionnaires included incomplete, unclear, illogical or inconsistent information. The author or a research assistant phoned the parents to clarify such answers and fill in missing information. When the parents could not be reached or when unclear or inconsistent information still remained, the BiLI-TAS research team held several detailed discussion sessions to resolve matters.
time of testing. One of these children (BiAra4-17) was seen again one year later when she was 4 years. Of the 100 children included in the cross-sectional study in Sweden, no child had a diagnosis of Developmental Language Disorder (DLD), a neuropsychiatric disorder or hearing impairment at the time of recruitment, according to the parents.107

The majority of the children lived in the greater Uppsala and Stockholm regions while some of children lived in the Västerås region. The participating children attended 53 different preschools and schools within these regions. All the 4-year-olds and 5-year olds attended preschool. In the 6-year-old group, apart from 3 children who still attended preschool, all children attended preschool class, förskoleklass, which is a mandatory year in between preschool and compulsory primary school in Sweden. In the 7-year-old group, 19 children attended first grade, and for the remaining five, no data was available whether they attended first grade or förskoleklass. General information about the participants in terms of number, gender, mean age and age range in each age group is provided in Table 5.1.

Table 5.1. Participants in the cross-sectional study in Sweden: number, gender, mean age and age range per age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Girls/boys</th>
<th>Mean age (years; months)</th>
<th>Age range (years; months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-year-olds</td>
<td>22</td>
<td>12/10</td>
<td>4;5</td>
<td>4;0–4;11</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>25</td>
<td>10/15</td>
<td>5;6</td>
<td>5;0–5;11</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>29</td>
<td>12/17</td>
<td>6;6</td>
<td>6;0–6;11</td>
</tr>
<tr>
<td>7-year-olds</td>
<td>24</td>
<td>16/8</td>
<td>7;7</td>
<td>7;1–7;11</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>50/50</td>
<td>6;1</td>
<td>4;0–7;11</td>
</tr>
</tbody>
</table>

Regarding the children’s place of birth, 55 children were born in Sweden, whereas 44 children were born in an Arabic-speaking country. One child (BiAra6-11) was born in an English-speaking country.

The Arabic-speaking population in Sweden is diverse, and this is reflected in the Arabic varieties spoken by the children participating in the study. Syrian and Palestinian varieties were the most commonly spoken, followed by

---

107 The parents of five children reported that they had been in contact with a speech-language pathologist (SLP), the reason usually being minor problems in the child’s pronunciation or the parents’ perception that their child was a late talker. The parents of one child (BiAra5-20) reported that they intended to contact an SLP in the future. It was eventually revealed to us that child BiAra5-20 was diagnosed with DLD around 1.5 years after data collection for the cross-sectional study. The child was not excluded from the study since she was not diagnosed with DLD at the time of testing and could complete all language tasks in both languages with comparable results to her peers. Furthermore, we cannot be sure that none of the other 99 children participating in the study was diagnosed with DLD later, therefore, it seemed unreasonable to exclude BiAra5-20 from the study just because we found out later on.
Iraqi, Lebanese and Egyptian. Table 5.2 gives an overview of the distribution of the Arabic varieties spoken by the children in the cross-sectional study per age group, as designated in the parental questionnaire.

Table 5.2. Overview of the Arabic varieties spoken by the children in each age group.

<table>
<thead>
<tr>
<th>Arabic Variety</th>
<th>4-year-olds</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>7-year-olds</th>
<th>Total (N = 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egyptian</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Iraqi</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Lebanese</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Palestinian</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>Syrian</td>
<td>7</td>
<td>12</td>
<td>15</td>
<td>10</td>
<td>43</td>
</tr>
</tbody>
</table>

Note. The percentage is the same as the number of participants (N = 100).

Some children were reported to speak a third language in the home in addition to Arabic and Swedish. The most common third language that the children spoke at home was English (N = 5). One child (BiAra7-11) was reported to speak Kurdish (Sorani) in addition to Arabic. Two children (BiAra4-08 and BiAra6-26) were also reported to speak Modern Standard Arabic (MSA) at home. Trilingual children or children who were reported to have a passive knowledge of a third language were not excluded from the cross-sectional study since multilingualism is part of the Arabic-Swedish-speaking population in Sweden. Furthermore, these children were able to complete all the tasks in the test battery in both languages.

5.1.2 Age of onset and rated language proficiency

Via the parental questionnaire, parents reported when their child started to receive regular exposure to Arabic and to Swedish, as well as their evaluation of the child’s proficiency in terms of comprehension and production in both languages.

108 The children speaking the Iraqi varieties spoke those of the Baghdad and Mosul regions.
109 Children who spoke Egyptian Arabic varieties were not excluded from the study. The experimenter had prepared additional scoring sheets for the Arabic testing session with children who spoke an Egyptian Arabic variety. In contrast to the encounter with the Sudanese-speaking children (who were excluded from the study), there appeared to be a mutual understanding between the experimenter and the Egyptian-speaking children.
110 In a few cases, one parent reported that s/he was born or had been raised in a different Arabic-speaking country than the partner or that the parents spoke different Arabic varieties from each other. These additional Arabic varieties spoken by the parents (other than that reported for the children) include Jordanian, Kuwaiti, Libyan, Moroccan and Saudi Arabian. For more information regarding children’s exposure to additional Arabic varieties, see Section 5.1.5.
111 Some children may also have been exposed to some Syriac/Aramaic, since they were recruited from associations and congregations where official ceremonies were conducted in Syriac/Aramaic, or where adults occasionally speak to each other in Syriac/Aramaic in addition to Arabic and Swedish. Surprisingly, however, no parent reported Syriac/Aramaic in the questionnaire.
Regarding the child’s regular exposure to Arabic, parents were asked to tick a box on a scale from birth to 7 years indicating during which year of life the child had started to receive regular exposure to each language. Ninety-eight children were reported to have had regular exposure to Arabic from birth. Only one child (BiAra5-25) started to have regular exposure to Arabic between the age of 1;0 and 2;0 and regular exposure to Swedish from birth. Children’s age of onset in Swedish, however, varied much more. Forty-eight children had regular exposure to Swedish before the age of 3;0, only 7 out of which had had regular exposure to Swedish from birth. Fifty-one children were first exposed to Swedish after the age of 3;0. Table 5.3 provides a summary of the age of onset for Arabic and for Swedish.

**Table 5.3.** Number of children with age of onset (AoO) of regular exposure for Arabic and for Swedish.

<table>
<thead>
<tr>
<th>Age of Onset</th>
<th>AoO in Arabic</th>
<th>AoO in Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between birth and age 1;0</td>
<td>98</td>
<td>7</td>
</tr>
<tr>
<td>Between 1;0 and 2;0</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Between 2;0 and 3;0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Between 3;0 and 4;0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>Between 4;0 and 5;0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Between 5;0 and 6;0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Between 6;0 and 7;0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Between 7;0 and 8;0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing information</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

*Note*. The percentage is the same as the number of participants (Total $N = 100$).

Twenty children had had less than 2 years of exposure to Swedish at the time of testing. These children were, however, immersed in the Swedish language via preschool or school and were able to complete all the tasks in Swedish in addition to Arabic. Children who recently moved to Sweden were not excluded since they are an integral part of the Arabic-Swedish-speaking population in Sweden. The BiLI-TAS research team chose not to exclude children who had had late AoO or short exposure length (in either language) because we wanted to explore the children’s test scores in relation to length of exposure. As long as a child could complete the tasks in both languages (active bilinguals), s/he was not excluded from the study.

Regarding the children’s language proficiency, parents were asked to estimate their child’s proficiency in comprehension and production in both Arabic and Swedish. The parents ticked one box out of five (on a five-point scale) ranging from ‘very good’, ‘good’, ‘so-so’, ‘poor’ to ‘very poor’. The majority of the parents reported that their children have a ‘very good’ or a ‘good’ pro-
iciency in both Arabic and Swedish comprehension and production. However, more parents estimated that their children had better proficiency (‘very good’) in Arabic than Swedish,\textsuperscript{112} as can be seen in Table 5.4.

**Table 5.4.** Number of children in each category of reported proficiency for comprehension and production of Arabic and Swedish, according to parents.

<table>
<thead>
<tr>
<th>Arabic comprehension</th>
<th>Arabic production</th>
<th>Swedish comprehension</th>
<th>Swedish production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good</td>
<td>69</td>
<td>54</td>
<td>45</td>
</tr>
<tr>
<td>Good</td>
<td>23</td>
<td>31</td>
<td>44</td>
</tr>
<tr>
<td>So-so</td>
<td>5</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Poor</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Very poor</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing information</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

5.1.3 Social background and language knowledge of the parents

The questionnaire also targeted the social and linguistic background of the parents. Parents were asked to specify their education level, their country of birth and where they grew up, their length of residence in Sweden, and their first language (L1). They were also asked about proficiency in their second language (Swedish) and whether they knew any other languages. This information was collected for both parents, except when a child was living with one parent only.

Socio-Economic status (SES) has often been reported to affect children’s language development. In the present study, parental education was used as a proxy for SES. Parents’ current occupations ranged over a wide range of fields. Regardless of family income, however, families residing in Sweden have access to public day-care and schools for their children.\textsuperscript{113} In the questionnaire, both parents were asked to report their highest level of education. The answers were coded according to the United Nations’ ISCED 2011 classification (UNESCO Institute for Statistics, 2012). Education level was thus ranked on the ISCED nine-level scale, ranging from 0 (early childhood education) to 8 (doctoral studies). The parents’ education levels varied considerably where all levels from 0 to 8 were present. Parental education was evenly distributed across the age groups. The mean level of education was 4.42 ($N = 91, SD = 2.06$) for Parent 1 and 4.16 ($N = 92, SD = 2.02$) for Parent 2.

\textsuperscript{112} The parents’ proficiency in a certain language may influence their evaluation of their child’s proficiency (Tuller, 2015).

\textsuperscript{113} No information regarding the household income was gathered as querying this information can be sensitive. Besides, an educated individual who is a recent newcomer to a country (as a refugee or otherwise) is not guaranteed to find a job instantly in that new country, regardless of previous work experience or education level.
mation regarding parental education was missing for 17 parents (for 5 children, information was missing for one of the parents, and for 6 children, information was missing from both parents).

The majority of the parents had been born in an Arabic-speaking country. Most of them had also grown up in an Arabic-speaking country. A few parents \((N = 8)\) had been born in an Arabic-speaking country but grew up in Sweden. Only one parent was born and raised in Sweden. For nine parents, information was missing for country of birth and country of upbringing. Table 5.5 shows the distribution of parents according to their country of birth and upbringing.

**Table 5.5.** Parents’ country of birth and upbringing.

<table>
<thead>
<tr>
<th></th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Born and brought up in an Arabic-speaking country</td>
<td>91</td>
<td>83</td>
</tr>
<tr>
<td>Born and brought up in Sweden</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Born in an Arabic-speaking country, brought up in Sweden</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Born in an Arabic-speaking country, brought up in another country</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Born in another country, brought up in an Arabic-speaking country</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Missing information</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

*Note.* The numbers reflect both \(N\) and the percentage of the total sample \((N = 100)\).

Regarding the parent’s first language (L1), the vast majority of the parents reported that they had Arabic as their first language. No parent reported that they had Swedish as their L1. One parent (of child BiAra7-11) reported that Arabic was not their first language, but did not indicate which language was their L1.\(^{114}\) No information was available regarding the parent’s first language for 7 parents in total. It could be cautiously assumed that the majority of the missing information is also Arabic since 98 households indicated that the parents only spoke Arabic with each other. Table 5.6 provides a summary of both parents’ L1.

**Table 5.6.** Parents’ first language.

<table>
<thead>
<tr>
<th></th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic as L1</td>
<td>98</td>
<td>94</td>
</tr>
<tr>
<td>Swedish as L1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other language as L1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Missing information</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

\(^{114}\) It may be assumed that this parent has Kurdish (Sorani) as L1 as s/he reported that s/he often spoke Kurdish along with Swedish with the child at home.
As for the parents’ length of residence in Sweden, the numbers varied greatly from 10 months being the shortest and 31 years being the longest. (These numbers do not include the one parent that indicated that s/he was born and raised in Sweden.) The mean length of stay for Parent 1 was 9.16 years ($N = 95$, $SD = 7.21$) and for Parent 2 it was 10.02 years ($N = 92$, $SD = 8.09$). No information was available regarding the parent’s length of stay in Sweden for 12 parents, where for 3 children, information was missing from both parents.

Parents were also asked to rate their own proficiency in Swedish. The results varied greatly, mirroring the large variation in residence length in Sweden. Almost one third of the parents ($N = 29$ for Parent 1, $N = 33$ for Parent 2) did not provide information regarding their language proficiency. Leaving this question blank may indicate low proficiency in Swedish.

Table 5.7. Distribution of parents’ proficiency in Swedish (self-assessed).

<table>
<thead>
<tr>
<th>Mother tongue</th>
<th>Fluent, almost native</th>
<th>Intermediate knowledge</th>
<th>Basic knowledge</th>
<th>No knowledge</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent 1</td>
<td>0</td>
<td>15</td>
<td>35</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Parent 2</td>
<td>0</td>
<td>13</td>
<td>35</td>
<td>17</td>
<td>2</td>
</tr>
</tbody>
</table>

The parents were further asked whether they had a knowledge of other languages (in addition to Arabic and Swedish). Almost all of the parents who indicated that they spoke an additional language to Arabic and Swedish noted that they had a knowledge of English ($N = 44$ for Parent 1, $N = 33$ for Parent 2). Other additional known languages included Aramaic/ Syriac, Armenian, Chaldean, Czech, French, Greek, Kurdish, Romanian, and Turkish.

5.1.4 Language use in the home

Language use in the home was explored through questions related to the language spoken between the parents, the parents’ language to the child, the child’s language to the parents and the child’s language to the siblings.

The language the parents speak with each other at home gives an idea of what language the child is exposed to (even if not directly addressed to) on a daily basis. The vast majority of the parents reported that they spoke almost only Arabic at home. No parental questionnaire reported that the parents spoke predominately Swedish with each other. The parents of three children reported that they spoke both Arabic and Swedish to each other. No information was available from four households regarding parental language use with each

---

115 Keep in mind that these estimations are based on the parents’ own view on their Swedish language proficiency, they might or might not reflect in reality how well they speak that language. Subjective estimations can either be too ambitious or very modest.

116 This question was left blank in many questionnaires; missing information was $N = 52$ for Parent 1 and $N = 66$ for Parent 2. Leaving out such information may indicate that the parents do not know any third language.
other.\textsuperscript{117} Table 5.8 summarizes the parents’ reported language use with each other.

**Table 5.8. Parents’ language use with each other.**

<table>
<thead>
<tr>
<th>Only Arabic</th>
<th>Arabic and Swedish</th>
<th>Only Swedish</th>
<th>Other</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>100</td>
</tr>
</tbody>
</table>

The language parents speak with the child on a daily basis provides a direct language input source. The parents were asked to specify on a five-point scale their language use with their child: ‘almost only Arabic’, ‘mostly Arabic’, ‘even’, ‘mostly Swedish’, ‘almost only Swedish’. The parents could also fill in an ‘other’ category if their language use did not match that of the scale. The majority of the parents reported that they spoke mainly (only or mostly) Arabic at home with their child. No parental questionnaire reported that the parents spoke predominately Swedish with the child. The parents of two children reported ‘other’, involving a third language.\textsuperscript{118} Table 5.9 summarizes the parents’ reported language use with their child.

**Table 5.9. Parents’ language use with child.**

<table>
<thead>
<tr>
<th>Both parents mainly Arabic</th>
<th>79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both parents mainly Swedish</td>
<td>0</td>
</tr>
<tr>
<td>1 parent mainly Arabic, 1 parent 50/50</td>
<td>8</td>
</tr>
<tr>
<td>1 parent mainly Arabic, 1 parent mainly Swedish</td>
<td>1</td>
</tr>
<tr>
<td>1 parent mainly Swedish, 1 parent mainly 50/50</td>
<td>0</td>
</tr>
<tr>
<td>Both parents 50/50</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
</tr>
<tr>
<td>Missing information</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

The parents were also asked to report which language(s) the child speaks to them. Table 5.10 shows the child’s language use with their parents. Similarly to the pattern of parental language use to the child, the majority of the children were reported to speak mainly (only or mostly) Arabic with their parents. The parents of eight children, however, reported that their children spoke mainly Swedish with them. This pattern differs from the parent-to-child language use, where no parent indicated that they spoke predominately Swedish with their child.

The parents of seven children reported that their child spoke mainly Arabic to one parent and equal amounts of Arabic and Swedish to the other parent.

\textsuperscript{117} Two parents (one parent of BiAra6-16 and one of BiAra7-10) barely reported any information about Parent 2. These children were thus assumed to live in single-parent households.

\textsuperscript{118} The parents of BiAra7-11 reported speaking Kurdish 60% and Swedish 40% (Parent 1) and ‘only Arabic’ (Parent 2) to the child. The parents of BiAra4-04 reported that they spoke Arabic and German (Parent 1) and ‘mostly Arabic’ (Parent 2).
Only one household reported that their child spoke mainly Arabic to one parent and mainly Swedish to the other. Thirteen children spoke both Arabic and Swedish evenly with both parents. One child was reported to speak mainly Swedish to one parent and equal amounts of Arabic and Swedish to the other parent. One child (BiAra7-11) was reported to speak Kurdish (Sorani) and Swedish to one parent, and mainly Arabic to the other. For one child, information regarding the language use with the parents was missing.

Table 5.10. Children’s language use with parents.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>To both parents mainly Arabic</td>
<td>68</td>
</tr>
<tr>
<td>To both parents mainly Swedish</td>
<td>8</td>
</tr>
<tr>
<td>To 1 parent mainly Arabic, with 1 parent 50/50</td>
<td>7</td>
</tr>
<tr>
<td>To 1 parent mainly Arabic, with 1 parent mainly Swedish</td>
<td>1</td>
</tr>
<tr>
<td>To 1 parent mainly Swedish, with 1 parent 50/50</td>
<td>1</td>
</tr>
<tr>
<td>To both parents 50/50</td>
<td>13</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>Missing information</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

As for the children’s language use with the sibling(s), the parents were asked to choose one option of five: ‘no sibling’, ‘mostly Arabic’, ‘both Arabic and Swedish’, ‘mostly Swedish’, or ‘other language’. The children’s pattern of language use with their siblings did not mirror that of their language use with their parents. Table 5.11 shows an overview of the children’s reported language use with their siblings. Six children were reported not to have any siblings. A little over one third of the children (N = 39) spoke mostly Arabic to their siblings whereas about another one third (N = 36) spoke both Arabic and Swedish. Eighteen children spoke mostly Swedish. The parents of four children reported ‘other’ language use between siblings, i.e. occasional use of English in addition to Swedish and/or Arabic. For 1 child, no information regarding language use with siblings was available.

Table 5.11. Children’s language use with sibling(s).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No sibling</td>
<td>Mostly Arabic</td>
</tr>
<tr>
<td>6</td>
<td>39</td>
</tr>
</tbody>
</table>

Parents were asked about their child’s birth order i.e. the number that indicated the order in which their child was born with respect to siblings. As can be seen

119 There were 10 sibling pairs in the data (i.e. for 10 families, 2 of their children participated in the study).
120 The parents of four children (BiAra5-06, BiAra6-17, BiAra7-17, BiAra7-23) reported that their child also occasionally spoke English with their sibling(s) in addition to either mostly Arabic (N = 2), mostly Swedish (N = 1) or both languages (N = 1). Since in none of the four cases English was reported as the only or predominant language spoken between the siblings, these cases were not coded as a separate ‘other’ category.
in Table 5.12, thirty-nine children are first-borns (do not have an older sibling),\textsuperscript{121} whereas sixty-one children have at least one older sibling.

**Table 5.12.** Children’s birth order amongst siblings.

<table>
<thead>
<tr>
<th>Birth Order</th>
<th>1(^{st}) Born</th>
<th>2(^{nd}) Born</th>
<th>3(^{rd}) Born</th>
<th>4(^{th}) Born</th>
<th>5(^{th}) Born</th>
<th>6(^{th}) Born</th>
<th>7(^{th}) Born</th>
<th>Missing Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^{st}) Born</td>
<td>39</td>
<td>31</td>
<td>18</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

For each language, parents were asked to specify on a 4-point scale (‘almost everyday’, ‘once/twice a week’, ‘twice a month’ or ‘(almost) never’) how often they read with their child at home. As can be seen in Table 5.13, the majority of the parents seem to read with their child at home at least once a week. However, note the relatively large amount of data missing on this question.

**Table 5.13.** Parent-and-child joint reading activity at home in Arabic and in Swedish.

<table>
<thead>
<tr>
<th>Language</th>
<th>Almost Everyday</th>
<th>Once/twice a Week</th>
<th>Twice a Month</th>
<th>Almost Never</th>
<th>Missing Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>14</td>
<td>39</td>
<td>13</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>Reading in Swedish</td>
<td>20</td>
<td>29</td>
<td>15</td>
<td>19</td>
<td>17</td>
</tr>
</tbody>
</table>

5.1.5 Home language use outside the home

Bilingual children often have the opportunity to speak their home language outside their home as well. The parents were asked to report their children’s language use with extended family, with friends and staff members at (pre)school, as well as with friends (and staff) during extracurricular activities (Table 5.14).\textsuperscript{122}

The majority of the children (\(N = 78\)) were reported to hear Arabic from extended family.\textsuperscript{123} About one fifth of the children (\(N = 21\)) were not reported to hear Arabic from extended family. For five children there was no data available. Only about one third (\(N = 30\)) of the children were reported to hear Arabic from their friends, whereas about two thirds (\(N = 71\)) did not have any Arabic-speaking friends. No data was reported for nine children. Staff at preschool and school who are in daily contact with the child may sometimes speak Arabic as well. The majority of the children were not reported to have contact with an Arabic-speaking staff member (\(N = 72\)), whereas one fourth

\textsuperscript{121} Six out of the 39 first-born children did not have any younger siblings.

\textsuperscript{122} Note that the parents were asked whether their child heard Arabic from extended family, friends and staff members in the form of a ‘multiple-choice question’ where they could mark one or several answers. Hence, not having selected an answer does not necessarily mean that the child does not hear Arabic, but might be related to the parent’s carelessness while filling in the question.

\textsuperscript{123} No information is available regarding how frequent or how long such encounters are.
of the children did have contact with an Arabic-speaking person at their (pre)school. Two children (BiAra4-07, BiAra5-06) had English-speaking staff at their preschool (both children went to the same preschool). No information was available for 3 children.

Table 5.14 Children hear Arabic from extended family, friends, and (pre)school staff.

<table>
<thead>
<tr>
<th></th>
<th>Does not hear Arabic</th>
<th>Hears Arabic</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended family</td>
<td>21</td>
<td>78</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Friends</td>
<td>70</td>
<td>29</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>(Pre)school Staff</td>
<td>72</td>
<td>25</td>
<td>3</td>
<td>100</td>
</tr>
</tbody>
</table>

Parents were asked to report whether their children heard other Arabic varieties than what the child spoke with her/his family. Two thirds of the children ($N = 62$) were not reported to hear any other variety of Arabic whereas about one third of the children ($N = 37$) were reported to hear other varieties of Arabic than the one they speak. When asked to specify where their child heard other Arabic varieties, parents reported close family members, (pre)school, Mother Tongue Instruction (MTI) class or on TV. These Arabic varieties included Egyptian, Iraqi, Jordanian, Lebanese, Libyan, Kuwaiti, Moroccan, Palestinian, Saudi Arabian and Syrian. Three parents indicated that their child heard another Arabic variety when they travel abroad when visiting extended family members.

When asked whether the children heard Modern Standard Arabic (MSA), the numbers were reversed. One third of the children ($N = 35$) were not reported to have had any exposure to MSA, whereas two thirds ($N = 64$) were reported to hear MSA frequently. Almost all the parents indicated that their children heard MSA from the TV or online video channels, whereas many also mentioned school (MTI classes) and books. Two parents reported that they occasionally spoke MSA at home with the child.\textsuperscript{124} Table 5.15 summarizes the children’s exposure to other Arabic varieties and to MSA according to parental report.

Table 5.15. Children exposed to other varieties of Arabic and to MSA.

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hears other varieties of Arabic</td>
<td>37</td>
<td>62</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Hears MSA</td>
<td>64</td>
<td>35</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>

\textsuperscript{124} One parent of BiAra4-08 reported to often speak MSA with the child at home. This child performed both narrative tasks (MAIN) in the Arabic session in MSA.
5.1.6 Additional language input sources in the home

To further investigate the quantity and quality of language input in both languages at home, parents were also asked to report how often they had performed the following activities with their child in the past month: telling stories, reading books, listening and singing to songs, and watching videos in Arabic and Swedish (Table 5.16). Although it is not possible to measure the richness of the language that is being provided in the books the children read or the songs they hear, these activities ensure that a certain amount of language input is provided.

**Table 5.16.** Additional language input sources in Arabic and Swedish.

<table>
<thead>
<tr>
<th>Language</th>
<th>Activity</th>
<th>Never</th>
<th>Twice a month</th>
<th>Once/twice a week</th>
<th>Almost every day</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>Tell stories</td>
<td>7</td>
<td>17</td>
<td>42</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Singing/listening to songs</td>
<td>10</td>
<td>12</td>
<td>22</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Watch TV/movies</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>62</td>
<td>11</td>
</tr>
<tr>
<td>Swedish</td>
<td>Tell stories</td>
<td>26</td>
<td>16</td>
<td>20</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Singing/listening to songs</td>
<td>15</td>
<td>14</td>
<td>22</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Watch TV/movies</td>
<td>15</td>
<td>7</td>
<td>12</td>
<td>53</td>
<td>13</td>
</tr>
</tbody>
</table>

In general, parents seem to tell stories in Swedish to a lower degree than in Arabic with children at home. Children learn more to sing in their home language with their parents, rather than in Swedish. Furthermore, the majority of the children tend to be exposed to language input everyday by watching videos (via TV or other electronic devices), even more so in Arabic than in Swedish. As can be seen in Table 5.16, many parents chose not to answer this question regarding these activities with their child. There is a lot of ‘missing information’ in both languages. It may be cautiously assumed that these activities are never or seldom performed at home by the parents and the child together. Another possible explanation for the large number of missing data could be related to the position of the question at the end of the extensive parental questionnaire.

5.1.7 Mother-tongue instruction

Municipalities in Sweden offer Mother Tongue Instruction (MTI) for children who speak a language other than Swedish in the home. Mother Tongue Instruction (MTI) classes usually (but not necessarily) take place after school hours, together with other children for about 40 minutes to 1 hour per week. Also, some institutions and congregations offer private classes of MTI. The
parents were asked to indicate whether their children attended such additional classes in Arabic, if so where, for how long and whether such MTI classes were attended together with other children.

Table 5.17 shows MTI class attendance and the type of MTI classes broken down by age group. All in all, around two-thirds of the children \((N = 65)\) were reported to attend some sort of MTI class (either via the municipality, via private initiative, or both). More older children (in the 6-and 7-year-old groups) attended municipality MTI classes, while more younger children (in the 4- and 5-year-old groups) did not. A reasonable explanation for this is that municipalities are nowadays no longer required to organize MTI instruction at preschools as they are in schools. Hence, more younger children (in the 4- and 5-year-old groups) seem to attend MTI via private initiatives.

### Table 5.17. Mother Tongue Instruction (MTI) attendance by type and age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Attends via municipality</th>
<th>Attends via private initiative</th>
<th>Attends both via municipality and private initiate</th>
<th>Does not attend MTI</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>21</td>
<td>6</td>
<td>31</td>
<td>4</td>
</tr>
</tbody>
</table>

#### 5.1.8 Estimated daily language exposure

Parents were asked to estimate the proportion of daily exposure their child receives in both languages. On a seven-point scale, parents indicated the relative proportion of exposure their child had to Arabic and Swedish during the day. Parents were asked the following question: ‘How often does your child hear both languages in their everyday life? Put a mark on the scale.’ Each of the seven options on the scale contained a percentage of exposure of Arabic and Swedish. The categories were: ‘Swedish 5% and Arabic 95%’, ‘Swedish 20% and Arabic 80%’, ‘Swedish 40% and Arabic 60%’, ‘Swedish 50% and Arabic 50%’, ‘Swedish 60% and Arabic 40%’, ‘Swedish 80% and Arabic 20%’, ‘Swedish 95% and Arabic 5%’. Parents could also indicate on an eighth point (named ‘other’) whether their child’s daily exposure pattern was different from the scale and for instance included a third language. Table 5.18 shows an overview of the distribution of daily language exposure by age group, as estimated by the parents.
Table 5.18. Estimated daily language exposure as per age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Swe 5%</th>
<th>Swe 20%</th>
<th>Swe 40%</th>
<th>Swe 50%</th>
<th>Swe 60%</th>
<th>Swe 80%</th>
<th>Swe 95%</th>
<th>Ara 95%</th>
<th>Ara 80%</th>
<th>Ara 60%</th>
<th>Ara 50%</th>
<th>Ara 40%</th>
<th>Ara 20%</th>
<th>Ara 5%</th>
<th>Other</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>13</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>11</td>
<td>18</td>
<td>37</td>
<td>22</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note. Swe = Swedish, Ara = Arabic.

About one third \((N = 37)\) of the children were reported to have had an even exposure \((50/50)\) to both languages per day. A little less than one fifth \((N = 18)\) were reported to have ‘Swedish 40%-Arabic 60%’, whereas a little more than one fifth \((N = 22)\) were reported to have ‘Swedish 60%-Arabic 40%’ exposure. Therefore, the majority of the children \((N = 77)\) have a relatively balanced exposure to Arabic and Swedish during the day. A total of 12 children were estimated to be exposed to more Arabic than Swedish \((80% \text{ or more exposure to Arabic})\). Most of these children are 4 or 5 years old and attend preschools where both Arabic and Swedish-speaking staff members surround the child. Only 8 children in total were reported to be exposed predominately to Swedish \((80\% \text{ or more exposure to Swedish})\). Two children were reported to have different language exposure patterns. One child (BiAra7-17) was reported to be exposed to English in addition to Arabic and Swedish, whereas another child (BiAra7-11) was reported to be exposed additionally to Kurdish (Sorani).

5.1.9 Attitudes towards language

The parents were asked about their children’s language preference and which language the parents consider as the most important for the child. Regarding the children’s language preference, slightly more than half of the children \((N = 52)\) were reported not to prefer one language over the other. One fifth of the children \((N = 20)\) were reported to prefer speaking Arabic, and about one fifth \((N = 22)\) were reported to prefer Swedish over Arabic. Two children (BiAra6-11 and BiAra7-17) were reported to prefer speaking English over the two languages. Information was missing for 3 children. Table 5.19 provides an overview of the children’s language preference.
Table 5.19. Child’s language preference.

<table>
<thead>
<tr>
<th>No preference</th>
<th>Prefers Arabic</th>
<th>Prefers Swedish</th>
<th>Prefers other</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>20</td>
<td>22</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

The parents were also asked to report their attitude towards which language they perceive as the most important for their child to learn. The majority of the parents ($N = 81$) reported that they believe that both Arabic and Swedish are equally important for their children. Some parents ($N = 12$) considered Arabic to be the more important language. Only two parents stressed the importance of the child learning Swedish over Arabic. Four parents reported that both Arabic and Swedish in addition to English are all important languages for the child to learn.

Table 5.20. Parents’ language attitude for their children.

<table>
<thead>
<tr>
<th>Both equally important</th>
<th>Arabic most important</th>
<th>Swedish most important</th>
<th>Other</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

5.1.10 Summary of the participants in the cross-sectional study in Sweden

A little less than half of the children ($N = 44$) were born in an Arabic-speaking country while the remaining children (except for 1 child) were born in Sweden. The children spoke various Arabic varieties, namely Syrian ($N = 43$), Palestinian ($N = 26$), Iraqi ($N = 17$), Lebanese ($N = 9$) and Egyptian ($N = 4$). Age of onset of Arabic was at birth for all children, except for 1 child where it was shortly after birth. Age of onset of Swedish varied greatly and ranged from birth to age 7. About one third of the children ($N = 38$) started to receive frequent input to Swedish before the age of 3. In addition to Arabic and Swedish, a few children were reported to additionally speak English, Kurdish (Sorani) or MSA at home. Sixty-one children had at least one older sibling.

The children’s parents were reported to have been born and raised in an Arabic-speaking country (except for 1 parent who was born and raised in Sweden). The length of residency in Sweden ranged from 10 months to 31 years (mean $= 9.5$ years). The parents’ SES (education) ranged from having received only a couple of years of elementary schooling to having achieved a doctoral degree (mean level of education was having finished secondary school education). All the parents’ mother tongue was Arabic (except for 1 parent).

As for languages spoken at home, in the majority of households ($N = 90$), only Arabic was spoken between the parents, while in very few households ($N = 3$) the parents spoke both Arabic and Swedish with each other. Similarly, in

---

125 The majority of these parents (10 out of 12) had children who were either 4 or 5 years old.
the majority of the households \((N=78)\), the parents spoke mainly Arabic with the child, while few \((N=9)\) spoke an even amount of Arabic and Swedish. Language use by the child to the parents and siblings differed, however. The majority of the children \((N=68)\) spoke mainly Arabic with their parents, 13 children spoke an even amount of both languages, and 8 children spoke mainly Swedish with their parents. To siblings, however, only 39 children spoke mostly Arabic, while almost the same number of children \((N=36)\) spoke both Arabic and Swedish, and 18 children spoke mostly Swedish. About half of the households reported that they performed parent-child joint book reading at least once a week in each language.

Home language use outside the home was mostly practised with the extended family \((N=78)\), while fewer children were reported to speak Arabic with their friends \((N=29)\) or with school staff \((N=25)\). About one third of the children \((N=37)\) heard other Arabic varieties than their own and about two thirds of the children \((N=64)\) frequently heard MSA. Out of the additional language exposure sources inside the home, watching television/movies/computer was the most frequent activity (almost every day) for both Arabic and Swedish.

As for mother tongue instruction (MTI), 75 children were reported to attend MTI either via the municipality, private initiative or both. The majority of these children were 6 or 7 years old and attended MTI via the municipality. The few 4-year-old and 5-year-old children who received MTI did so mainly via private initiatives.

The children’s daily language exposure patterns broke down into three major groups where each group was approximately one third of the sample: 37 children were reported to receive equal amount of language input in both languages, 30 children were exposed the majority \((60\% \text{ or more})\) of their day to Arabic, and 30 children were exposed to the majority \((60\% \text{ or more})\) of their day to Swedish.

Finally, regarding language attitude, about half of the children \((N=52)\) were reported not to have any preference towards any of the languages, while one fifth of the children \((N=20)\) preferred Arabic and about another one fifth \((N=22)\) preferred Swedish. Two children were reported to prefer English over Arabic and Swedish. The parents’ attitudes were somewhat different, where the majority \((N=81)\) reported that both languages are important, and 12 parents reported that they would prefer their child to learn Arabic over Swedish. Only 2 parents indicated a language preference for their children of Swedish over Arabic, and 4 parents added English to their language preference in addition to Arabic and Swedish.
5.2 Participants of the longitudinal study in Sweden

5.2.1 General information

A subgroup of 10 4-year-old children from the cross-sectional study in Sweden participated in the follow-up study approximately two years (23-28 months) after the first encounter when they were 6 years old. One child (BiAra4-08) was 7 years at the follow-up data collection.

The aim was to compare the language performance of children at age 4 with their performance at age 6 (while using the same tasks). Table 5.21 shows an overview of the participants in regard to their age at Time 1 (T1, the cross-sectional session), age at Time 2 (T2, the follow-up session), time gap in between the two sessions (between Time 1 and Time 2) as well as the type of education (class) they were attending at T2. At the second occasion (T2), two children were reported to attend grade 1 of primary school, whereas the remaining children (N = 8) attended pre-school class (förskoleklass).

Table 5.21. Overview of the participants (N = 10) in the longitudinal part of the study.

<table>
<thead>
<tr>
<th>Child</th>
<th>Age at T1 (year;month)</th>
<th>Time between T1 and T2 (in months)</th>
<th>Age at T2 (year;month)</th>
<th>Education type at T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiAra4-02</td>
<td>4;10</td>
<td>24</td>
<td>6;11</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-05</td>
<td>4;4</td>
<td>23</td>
<td>6;4</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-06</td>
<td>4;2</td>
<td>27</td>
<td>6;5</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-08</td>
<td>4;11</td>
<td>28</td>
<td>7;4</td>
<td>Grade 1</td>
</tr>
<tr>
<td>BiAra4-13</td>
<td>4;10</td>
<td>27</td>
<td>6;3</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-14</td>
<td>4;3</td>
<td>25</td>
<td>6;4</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-15</td>
<td>4;4</td>
<td>25</td>
<td>6;5</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-16</td>
<td>4;4</td>
<td>23</td>
<td>6;4</td>
<td>Pre-school class</td>
</tr>
<tr>
<td>BiAra4-18</td>
<td>4;6</td>
<td>28</td>
<td>6;11</td>
<td>Grade 1</td>
</tr>
<tr>
<td>BiAra4-24</td>
<td>4;10</td>
<td>23</td>
<td>6;10</td>
<td>Pre-school class</td>
</tr>
</tbody>
</table>

Note. T1= Time 1, T2= Time 2.

Table 5.22 details the Arabic varieties spoken by the children in the longitudinal study.

Table 5.22. Overview of the Arabic varieties spoken by the children in the longitudinal study.

<table>
<thead>
<tr>
<th>Arabic Variety</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egyptian</td>
<td>1</td>
</tr>
<tr>
<td>Iraqi</td>
<td>0</td>
</tr>
<tr>
<td>Lebanese</td>
<td>2</td>
</tr>
<tr>
<td>Palestinian</td>
<td>3</td>
</tr>
<tr>
<td>Syrian</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10</td>
</tr>
</tbody>
</table>

Note: Child BiAra4-08 was already 4;11 (almost 5-years-old) when tested in the cross-sectional study. It was impossible to test her at exactly 24 months later due to the national pandemic restrictions.
Similarly to the cross-sectional study, speakers of the Syrian and the Palestinian varieties of Arabic predominated in the longitudinal study. No Iraqi-speaking children participated in the longitudinal study (although Iraqi was the third most spoken Arabic variety in the cross-sectional study). Two families reported that additional Arabic varieties were spoken at home, namely Jordanian and Moroccan respectively.

As for the place of birth of the children in the longitudinal study, the majority of children (N = 8) were born in Sweden whereas two children were born in Arabic-speaking countries. Recall that almost half of the children in the cross-sectional study were born in Sweden (N = 45), one child was born in an English-speaking country and the remaining children (N = 54) were born in an Arabic-speaking country. Hence, the proportion of Sweden-born children in the longitudinal sample is larger than in the cross-sectional one.

All the children’s age of onset (AoO) in Arabic was at birth whereas the AoO of Swedish varied greatly, just as in the larger cross-sectional sample. Seven children who participated in the longitudinal study had had two years or less of regular exposure to Swedish at the time of the first testing. Recall that only one fifth (N = 20) of all the children who participated in the cross-sectional study had had less than two years of exposure to Swedish at the time of testing. Hence, the longitudinal sample does not fully represent the cross-sectional sample in terms of AoO of Swedish. Table 5.23. shows an overview of the AoO of both languages for the children participating in the longitudinal study.

**Table 5.23.** Age of onset (AoO) of regular exposure for Arabic and for Swedish for the children in the longitudinal study.

<table>
<thead>
<tr>
<th>AoO in Arabic</th>
<th>AoO in Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between birth and age 1;0</td>
<td>10</td>
</tr>
<tr>
<td>Between 1;0 and 2;0</td>
<td>0</td>
</tr>
<tr>
<td>Between 2;0 and 3;0</td>
<td>0</td>
</tr>
<tr>
<td>Between 3;0 and 4;0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
</tr>
</tbody>
</table>

As for the parents’ education, the levels varied between 1 (primary education) and 7 (Master’s degree) on the ISCED scale. The mean level of education was 4.55 (N = 9, SD = 1.94) for Parent 1 and 4.00 (N = 9, SD = 1.88) for Parent 2. For 1 child, both parents’ education information was missing. The distribution and mean education level resembles the distribution of the children in the cross-sectional study (less the fact that there was more missing data in the cross-sectional study).

---

127 There were only two Iraqi-speaking children amongst the 4-year-old children in the cross-sectional study. One family was not re-contacted due to transportation restrictions during the pandemic, the other family declined participation because of pandemic safety issues.
5.2.2 Language input and use in and outside the home

Information regarding language input and language use at home was gathered via two parental questionnaires, at age 4 and at age 6. Table 5.24 shows the parents’ language use with the child and with each other as reported at these two different times.

Table 5.24. Parents’ language use with the child and with each other at age 4 and age 6.

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
<th>Parent 1 to child</th>
<th>Parent 2 to child</th>
<th>Parent to Parent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>At age 4</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-02</td>
<td></td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-05</td>
<td></td>
<td>Mostly Arabic</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>BiAra4-06</td>
<td></td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-08</td>
<td></td>
<td>Mostly Arabic</td>
<td>Only Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-13</td>
<td></td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-14</td>
<td></td>
<td>Mostly Arabic</td>
<td>Only Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-15</td>
<td></td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-16</td>
<td></td>
<td>Mostly Arabic</td>
<td>Mostly Arabic</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-18</td>
<td></td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic</td>
</tr>
<tr>
<td>BiAra4-24</td>
<td></td>
<td>Mostly Arabic</td>
<td>Mostly Arabic</td>
<td>Arabic</td>
</tr>
</tbody>
</table>

Note. N/A = Not applicable.

Although all parents reported that they spoke solely Arabic with each other, their language use towards the child showed a different pattern. The majority of the parents reported that they spoke ‘only or mostly Arabic’ to the child, while others reported that they spoke a mixture of Arabic and Swedish. No parent reported that they spoke predominately Swedish to her/his child.

Language use as reported at the two data collection times (T1 and T2) had changed for most families. Four families (eight parents) reported that they had started to speak more Arabic (only or mostly Arabic) at home in comparison to two years back. Five parents reported that they had shifted to speak more Swedish with their child at age 6. For instance, in the parental interview, the parent of BiAra4-02 and the parent of BiAra4-06 mentioned that they had

---

128 The parents of child BiAra4-05 was missing at T1. Information regarding parental language use with each other is thus not applicable (N/A). Also, data was not fully available for parent 2.
started to speak more Swedish (than before) to their child since they had experienced that the child had difficulty in understanding everything that was spoken to them in Arabic. Six parents reported that they had not much changed their language use towards their child.

As for the child’s language use at home, Table 5.25 shows the child’s language use with each of the parents and with the siblings as reported at age 4 (T1) and age 6 (T2).

Table 5.25. Child’s language use with parent and sibling(s) at age 4 and age 6.

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
<th>Child to Parent 1</th>
<th>Child to Parent 2</th>
<th>Child to sibling(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiAra4-02</td>
<td>At age 4</td>
<td>Mostly Swedish</td>
<td>Mostly Swedish</td>
<td>Arabic &amp; Swedish</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
</tr>
<tr>
<td>BiAra4-05</td>
<td>At age 4</td>
<td>Mostly Arabic</td>
<td>N/A</td>
<td>Only Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Mostly Arabic</td>
<td>N/A</td>
<td>Arabic &amp; Swedish</td>
</tr>
<tr>
<td>BiAra4-06</td>
<td>At age 4</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Arabic &amp; Swedish</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Arabic &amp; Swedish</td>
<td>Mostly Arabic</td>
<td>Arabic &amp; Swedish</td>
</tr>
<tr>
<td>BiAra4-08</td>
<td>At age 4</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6 (7)</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td>BiAra4-13</td>
<td>At age 4</td>
<td>Mostly Arabic</td>
<td>Mostly Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Mostly Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td>BiAra4-14</td>
<td>At age 4</td>
<td>Arabic &amp; Swedish</td>
<td>Only Arabic</td>
<td>Only Swedish</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Mostly Arabic</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
</tr>
<tr>
<td>BiAra4-15</td>
<td>At age 4</td>
<td>Arabic &amp; Swedish</td>
<td>Arabic &amp; Swedish</td>
<td>Only Swedish</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Mostly Arabic</td>
<td>Mostly Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td>BiAra4-16</td>
<td>At age 4</td>
<td>Mostly Arabic</td>
<td>Mostly Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td>BiAra4-18</td>
<td>At age 4</td>
<td>Mostly Arabic</td>
<td>Missing</td>
<td>Only Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td>BiAra4-24</td>
<td>At age 4</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
<td>Only Arabic</td>
</tr>
</tbody>
</table>

Note: N/A= Not applicable.

Not surprisingly, the child’s language use towards the parents in general seemed to mirror that of the parents with the child. Most of the children spoke predominately Arabic and sometimes Swedish to their parents.

The answers provided in the questionnaire were in accordance with the parental interview for all of the parents, except for one. When conducting the oral interview with the parent of BiAra4-14, the answers shifted more towards Swedish. The child was reported to speak mostly Swedish to parents, even when spoken to in Arabic. According to the parents, the child understood when addressed in Arabic, but chose to reply in Swedish.

With regard to the child’s language use with the sibling(s), the answers differed somewhat from the language use with the parents. According to the parental questionnaire, a little more than half of the children prefer to speak ‘only Arabic’ \(N = 6\) to their siblings whereas the remaining children \(N = 4\) were reported to speak ‘both Arabic and Swedish’ with siblings. However,
when this same topic was discussed in the informal interview, it became evident that language use between siblings had shifted towards more Swedish. Four families (BiAra4-05, BiAra4-06, BiAra4-14, BiAra4-15) stressed in the oral parental interview that their children in fact predominantly spoke Swedish with their sibling(s) at home. Some of these parents mentioned that they greatly encouraged their children to speak Arabic with each other, but that the children often refuse to do so and continue in Swedish. The parent of BiAra4-13 reported that her child predominately speaks Arabic to her sibling but added that they would occasionally code-switch to Swedish. The parent of child BiAra4-16, whose child was reported to speak ‘only Arabic’ in the parental questionnaire, added in the interview that the child occasionally spoke a few words in Swedish and English to the sibling. Similarly, children BiAra4-02 and BiAra4-05 were also reported in the interview to speak some English at home with the sibling(s) in addition to Arabic and Swedish. Both BiAra4-14 and BiAra4-15’s parents mentioned that their children were starting to show an interest in speaking English since they have started to learn it at school and are being more exposed to it while using their computers at home.

This shift towards speaking more Swedish with the siblings is in line with several studies that show that children tend to speak the majority language more often with their siblings than with their parents (e.g. de Houwer 2009; Bridges & Hoff, 2014). The parents of three children (BiAra4-05, BiAra4-06, BiAra4-16) also expressed that their children had started to speak more Swedish at home compared to before. This was not the case though for three other children (BiAra4-08, BiAra4-18 and BiAra4-24) whose parents reported (in both questionnaire and interview) that their child still spoke ‘only or mostly Arabic’ with their sibling(s) and in general at home.

As for the language used with grandparents and relatives, all of the parents indicated that their children spoke merely Arabic with them. The relatives were often contacted via telephone or online video/audio call, since they did not live in Sweden. Several parents indicated that this was their main motivation to teach their children Arabic, mainly to keep in contact with their relatives. It seems that when children speak with their extended family, they are more or less forced to use their mother tongue since the relatives do not understand Swedish. At least this holds true for the older generation, namely the children’s grandparents. The parent of BiAra4-02 pointed out in the interview that her child had occasionally started to speak English (rather than Arabic) with non-Swedish speaking cousins and uncles/aunts.

5.2.3 Rated language proficiency

The parents were asked to estimate their child’s proficiency in Arabic and Swedish as well as which of the two languages they believed the child spoke the best. Table 5.26 shows these estimates as reported by the parents at both T1 (age 4) and T2 (age 6). For the estimation of the language proficiency, the
parents chose one option on a five-point scale: ‘very good’, ‘good’, ‘so-so’, ‘poor’ and ‘very poor’.

Table 5.26. Child’s best language and language proficiency as reported in the parental questionnaire at age 4 and age 6.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BiAra4-02</td>
<td>At age 4</td>
<td>Swedish</td>
<td>Good</td>
<td>Good</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Both equal</td>
<td>So-so</td>
<td>So-so</td>
<td>Good</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-05</td>
<td>At age 4</td>
<td>Arabic</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
<td>So-so</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Both equal</td>
<td>So-so</td>
<td>So-so</td>
<td>Very good</td>
<td>Good</td>
</tr>
<tr>
<td>BiAra4-06</td>
<td>At age 4</td>
<td>Both equal</td>
<td>Very good</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Both equal</td>
<td>So-so</td>
<td>So-so</td>
<td>Good</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-08</td>
<td>At age 4</td>
<td>Arabic</td>
<td>Very good</td>
<td>Good</td>
<td>So-so</td>
<td>So-so</td>
</tr>
<tr>
<td></td>
<td>At age 6 (7)</td>
<td>Arabic</td>
<td>Arabic</td>
<td>Good</td>
<td>So-so</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-13</td>
<td>At age 4</td>
<td>Arabic</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Arabic</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-14</td>
<td>At age 4</td>
<td>Both equal</td>
<td>Very good</td>
<td>Good</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Swedish</td>
<td>So-so</td>
<td>Good</td>
<td>Very good</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-15</td>
<td>At age 4</td>
<td>Both equal</td>
<td>Very good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Both equal</td>
<td>Very good</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
</tr>
<tr>
<td>BiAra4-16</td>
<td>At age 4</td>
<td>Arabic</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
<td>So-so</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Arabic</td>
<td>So-so</td>
<td>So-so</td>
<td>So-so</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-18</td>
<td>At age 4</td>
<td>Both equal</td>
<td>Very good</td>
<td>Very good</td>
<td>Very good</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Arabic</td>
<td>Very good</td>
<td>Very good</td>
<td>Very good</td>
<td>So-so</td>
</tr>
<tr>
<td>BiAra4-24</td>
<td>At age 4</td>
<td>Arabic</td>
<td>Very good</td>
<td>Missing</td>
<td>Missing</td>
<td>Missing</td>
</tr>
<tr>
<td></td>
<td>At age 6</td>
<td>Both equal</td>
<td>Very good</td>
<td>Very good</td>
<td>Missing</td>
<td>Missing</td>
</tr>
</tbody>
</table>

*Note. comp. = comprehension, prod. = production. Lang. = language.*

The parents’ answers regarding their child’s best language varied across the children. Six parents indicated no change in the child’s best language from age 4 to age 6. The four remaining parents experienced some change in language dominance. As for the children’s proficiency in each language, many of the parents indicated no change in how they perceive their child’s proficiency in Arabic and Swedish comprehension and production.
5.2.4 Language input at school and MTI

Originally, the follow-up Swedish data collection session was planned to take place at school (where the child mostly speaks Swedish), and the Arabic data collection session would take place at home (where the child is more likely to speak Arabic).

At school, the experimenter would observe the school environment (whether the school was predominantly monolingual or multicultural, the language(s) spoken by the children and the staff, and the proficiency of that language). At home, observations would be related to the language(s) the parent(s) and the child(ren) spoke with each other as well as the language of the television, radio or other media during the visit. However, due to repeated cancellations and rescheduling to comply with the pandemic restrictions, school and/or home visits were not always possible. Some parents preferred that both sessions were conducted at school ($N = 5$), whereas others insisted that we were at home (or in a nearby garden) at both times ($N = 4$). For one child, both sessions were done in a congregation meeting room ($N = 1$), thus neither at school nor at home. As a result, the originally planned school and home observations could not be carried out.

By interviewing the child and the parents, a more general idea was formed regarding the children’s language input at school. The parents of seven children reported that their child had at least one Arabic-speaking staff member at school. The majority of the children ($N = 7$) also mentioned that they had at least one Arabic-speaking peer in class. All children were reported to attend monolingual schools where Swedish was the language of instruction. However, the parents of three children mentioned that there are several Arabic-speaking staff members at their children’s school (the three children attend the same school). Only one child (BiAra4-06) attended a school where English was a language of instruction in addition to Swedish.

Nine children were reported to attend Mother Tongue Instruction (MTI) provided by the municipality at age 6. The parent of one child (BiAra4-02) had reported MTI attendance in the questionnaire but mentioned in the parental interview that the child had quit attending MTI classes due to the change of Arabic teacher from one that spoke the same Arabic variety as the child to a teacher who spoke a more distant Arabic variety. Three children (BiAra4-16, BiAra4-18, BiAra2-24) were reported to additionally attend private Arabic classes, making their weekly Arabic language classes range from 3-4 hours. One child (BiAra4-08) did not attend any MTI but was reported to have been taught Arabic at home. These parents also indicated (at both T1 and T2) that they often spoke MSA with the child. In the parental interview, the parents

129 The parent of BiAra4-13 indicated that her child, in addition to attending MTI classes provided by the municipality, had initially also attended private Arabic classes during the weekend (totally 2 hours/week). However, due to the pandemic, the private Arabic language classes were put on hold.
added that they often read Arabic (MSA) books (about 3 times a week) with the child at home and borrowed new books constantly from the library.

5.2.5 Estimated daily language exposure
The parents were asked to estimate their child’s daily language exposure. The parents of four children (BiAra4-15, BiAra4-16, BiAra4-18, BiAra4-24) estimated that their child heard more Arabic than Swedish daily. One parent (BiAra4-15) added that the child also heard a certain amount of English. The parents of three children (BiAra4-02, BiAra4-05, BiAra4-14) estimated that their child had an even exposure to both languages, while the parents of two children (BiAra4-06, BiAra4-13) estimated that their child was exposed more to Swedish than Arabic throughout the day. Recall that BiAra4-05 was reported to attend a school where instruction was at times given in English in addition to Swedish. Data was missing for one child (BiAra4-08).

However, as previously mentioned (Section 5.2.2), during the parental interview, the parents stated that their children started to speak more and more Swedish at home. English is also being introduced. Hence, a certain language shift is occurring in the children’s daily life as they grow older.

5.2.6 Parents’ and teachers’ attitudes towards the home language
All the interviewed parents stated that they wanted their child to learn Arabic. The main reason was keeping contact with Arabic-speaking members of the extended family that reside outside of Sweden. Another reason mentioned by a couple of parents was religious affiliation. Many of the parents expressed that society will teach their child Swedish eventually, but that teaching Arabic was their role as parents.

When asked about the (pre)school’s attitude towards speaking Arabic at home, about half the families (N = 6) answered that they were explicitly encouraged by their child’s teacher to speak Arabic with the child. Only one parent (of child BiAra4-15) reported that she was advised to speak Swedish at home with her child since the child had difficulty in understanding the teacher when she started to attend pre-school class. All the parents, however, gave the same answer concerning what advice they would give to other bilingual families, namely to speak the home language at home.

5.2.7 Summary of the participants in the longitudinal study in Sweden
Information about the children who participated in the longitudinal study (N = 10) was gathered via a parental questionnaire in addition to a parental interview and an informal interview with the child. Similarly to the cross-sectional
study, the majority of the children spoke the Syrian and Palestinian Arabic varieties and were exposed to Arabic from birth. The mean level of SES (parental education) was almost identical in the two studies. The parents spoke Arabic to each other and no parent spoke predominately Swedish to the child.

The longitudinal and cross-sectional studies did not match in terms of MTI attendance. All children in the follow-up study were reported to attend MTI while only about three-fourths of the 6-year-olds in the cross-sectional study were reported to do so. Also, in regard to age of onset of Swedish, proportionally more children in the longitudinal study had started to receive frequent Swedish input before the age of 3 than the children in the cross-sectional study.

The majority of the children ($N = 7$) mentioned that they had (at least) one Arabic-speaking friend/classmate and that they knew an Arabic-speaking school staff member. This number is proportionally high in comparison to the cross-sectional study where less than one third of the children were reported in the questionnaire to have an Arabic-speaking friend. The majority of the parents in the longitudinal study spoke mostly/only in Arabic to their children and no one reported that they spoke mostly in Swedish. In the questionnaire answers, the children’s language choice to the parents mirrored that of the parents’ language to them (i.e. mostly Arabic); however, in the informal parental interview, several parents mentioned that their children spoke more and more Swedish with them and with their siblings. The informal interview also revealed that some children had started to speak English with siblings and extended family members. Conducting the informal oral interview gave a greater understanding of the complexity of the language use situation at home.

5.3 Participants of the cross-sectional study in Lebanon

5.3.1 General information

In Lebanon, 100 Arabic-speaking bilingual children were recruited to take part in the Lebanese cross-sectional study.\textsuperscript{130} The children were aged between 4;0–8;1. All children were reported to have (Lebanese) Arabic\textsuperscript{131} as their home

\textsuperscript{130} Six additional children (two 3-year-olds, three 4-year-olds, and one 5-year-old) were also tested yet excluded shortly after from the study. Two children were excluded since they had not turned 4 years of age at the time of testing. The remaining children’s sessions were terminated by the experimenter in mid testing. One child was very unfocused and not able to concentrate on performing the tasks. Two children were excluded since they did not understand the instructions of the tasks in Arabic. These children were reported to have had received intensive input in French or English, rather than in Arabic at home. One child attended a preschool where Arabic and French were the main languages of instruction, however, the child was reported to only speak Armenian at home with her parents. At preschool, that child was reported to remain quiet most of the time and not able to actively participate in discussions and activities.

\textsuperscript{131} One child (BiLeb4-01) was reported to speak both Lebanese and Palestinian varieties. One other child (BiLeb5-19) was reported to speak Lebanese and Syrian.
language, and to speak either English or French since an early age. None of the participating children in the Lebanese cross-sectional study had any known diagnosis of DLD, neuropsychiatric disorders or hearing difficulties. General information about the children in terms of number, gender, mean age and age range is provided in Table 5.27.

Table 5.27. Participants of the cross-sectional study in Lebanon: number, gender, mean age and age range per age group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>4-year-olds</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>7-year-olds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>18</td>
<td>29</td>
<td>31</td>
<td>22</td>
<td>100</td>
</tr>
<tr>
<td>Girls/boys</td>
<td>12/6</td>
<td>13/16</td>
<td>15/16</td>
<td>9/13</td>
<td>49/51</td>
</tr>
<tr>
<td>Mean age (years; months)</td>
<td>4;5</td>
<td>5;5</td>
<td>6;6</td>
<td>7;5</td>
<td>6;0</td>
</tr>
<tr>
<td>Age range (years; months)</td>
<td>4;0–4;11</td>
<td>5;0–5;10</td>
<td>6;1–6;11</td>
<td>7;0–8;1</td>
<td>4;0–8;1</td>
</tr>
</tbody>
</table>

Note: Questionnaire data was not available for 9 of the 100 children. Basic information, such as age and gender, was available via the consent form.

Two children (BiLeb7-16 and BiLeb7-20) had recently turned 8 years. These children were included in the study since they were very close to the targeted age range and attended the same grade as the other recruited 7-year-olds.

The children participating in the Lebanese cross-sectional study lived mainly in greater Beirut, Maten and Kiserwan districts. A few children were living in Byblos or Batroun districts. The children attended 22 different (private) schools and preschools within these regions.

As for the children’s place of birth, 87 children were reported to have been born in Lebanon, and 4 children (BiLeb4-09, BiLeb5-09, BiLeb5-12, BiLeb7-17) in a north American country. Information for 9 children was missing.

5.3.2 Age of onset and rated language proficiency

In the parental questionnaire, the parents were asked to indicate when their child had started to receive regular exposure in Arabic and in a second language (English or French) and estimate the child’s proficiency in terms of comprehension and production in both languages.

Parents were asked to tick a box indicating during which year of life their child started to receive regular exposure to each language. Eighty children had received regular exposure to Arabic before the age of 3, the majority of whom

132 Three parents reported that they had contacted an SLP, two of them (BiLeb6-20, BiLeb6-22) regarding their child’s past or present stuttering behaviour, and one (BiLeb6-13) out of concern for their child’s apparent language delay. Two children were reported to have had hearing problems in the past, however, they were reported to have normal hearing at the time of testing.
\((N = 72)\) had regular exposure to Arabic from birth. As for the second language, seventy-seven children had received regular exposure to English or French before the age of 3 and about half of them \((N = 38)\) had received regular input to a second language from birth. All in all, twenty-four children were reported to have had regular input in both languages (Arabic and English or French) from birth. Table 5.28 provides a summary of the age of onset for Arabic and for the second language.

**Table 5.28.** Children’s age of onset (AoO) for both Arabic and a second language (English or French).

<table>
<thead>
<tr>
<th>AoO in Arabic</th>
<th>AoO in a second language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between birth and age 1;0</td>
<td>Between 1;0 and 2;0</td>
</tr>
<tr>
<td>72</td>
<td>6</td>
</tr>
<tr>
<td>Between 2;0 and 3;0</td>
<td>Between 3;0 and 4;0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Between 4;0 and 5;0</td>
<td>Between 5;0 and 6;0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Between 6;0 and 7;0</td>
<td>Between 7;0 and 8;0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing information</td>
<td>Missing information</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

Note. The percentage is the same as the number of participants (Total \(N = 100\)).

Similarly to the cross-sectional study in Sweden, the majority of the children participating in the cross-sectional study in Lebanon were exposed to Arabic from before the age of 3.\(^{133}\) The groups differ in the reported AoO for the second language. While only 48% of the children in the Swedish cross-sectional study were exposed to Swedish before the age of 3, 77% of the children in Lebanon were exposed to a second language before the age of 3.\(^{134}\)

As for the children’s language proficiency, parents were asked to estimate their children’s proficiency in comprehension and production on a five-point scale ranging from ‘very good’, ‘good’, ‘so-so’, ‘poor’ to ‘very poor’. The majority of the parents reported that their children have a ‘very good’ or a ‘good’ proficiency in both Arabic comprehension and production. The estimation that Arabic comprehension was higher than Arabic production was slightly higher, as can be seen Table 5.29.\(^{135}\)

\(^{133}\) While almost all children \((N = 98)\) in the cross-sectional study in Sweden were exposed to Arabic from birth, only seventy-two children \((N = 72)\) were reported the same in Lebanon. Keep in mind though the high number of missing data \((N = 12)\) in the Lebanese cross-sectional study.\(^{134}\) As previously mentioned, children in Lebanon are often additionally exposed to a third (or fourth) language. The language combination is mostly Arabic, English and French, where one of the latter two is the language used in (pre)school for education and the other can be often heard in the media or from acquaintances.\(^{135}\) The parents were also asked to estimate their children’s proficiency in Arabic and in English and French. Only the Arabic comprehension and production proficiency estimates are reported here.
Table 5.29. Category of reported proficiency for comprehension and production of Arabic according to parents.

<table>
<thead>
<tr>
<th>Arabic comprehension</th>
<th>Arabic production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good</td>
<td>59</td>
</tr>
<tr>
<td>Good</td>
<td>24</td>
</tr>
<tr>
<td>So-so</td>
<td>7</td>
</tr>
<tr>
<td>Poor</td>
<td>0</td>
</tr>
<tr>
<td>Very poor</td>
<td>0</td>
</tr>
<tr>
<td>Missing information</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

The Arabic comprehension and production proficiency estimates in both cross-sectional studies (in Sweden and in Lebanon) are quite similar; the majority of both parental groups estimate that their children have a high level of comprehension and production in Arabic.

5.3.3 Social background and language knowledge of the parents

The questionnaire also targeted the social and linguistic background of the parents. The parents were asked to specify their highest education level. The answers were coded according to the United Nations’ ISCED 2011 nine-level classification (UNESCO Institute for Statistics, 2012) ranging from 0 (early childhood education) to 8 (doctoral studies). This served as a proxy for SES. The parents’ education levels varied between 2 (lower secondary education, max 9 years) and 8 (doctoral studies). Parental education was evenly distributed across the age groups. The mean level of education was 5.57 ($N = 77$, $SD = 1.4$) for Parent 1 and 5.72 ($N = 67$, $SD = 1.2$) for Parent 2. No information regarding parental education was available for 56 parents. For 22 children, information was missing from both parents. For 7 children, information was missing for Parent 2. It is believed that these families were of a single household since no information was provided about Parent 2. The mean level of parental education in the Lebanese cross-sectional study was higher than that of the parents in the Swedish cross-sectional study. However, in Lebanon, there was a much higher proportion of missing information on parental education.

The parents were also asked about their country of birth and where they grew up, their length of residence in Lebanon and whether Arabic was their first language (L1). As can be seen in Table 5.30, the majority of parents (Parent 1 $N = 83$, Parent 2 $N = 70$) were born and raised in Lebanon.

---

136 The parents were also asked about their proficiency in their second language (L2) and whether they knew any other languages. These numbers will not be reported here.
Table 5.30. Distribution of parents’ country of birth and where they grew up.

<table>
<thead>
<tr>
<th></th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Born and brought up in Lebanon</td>
<td>83</td>
<td>70</td>
</tr>
<tr>
<td>Born elsewhere and brought up in Lebanon</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Born in Lebanon but brought up elsewhere</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Born and brought up in an Arabic-speaking country</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Single parent household (no Parent 2 data)</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Missing information</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Regarding the parents’ first language (L1), the large majority of the parents reported that Arabic was their first language. Only two parents (in fact, one parent of the siblings BiLeb4-14 and 6-24) was reported to speak Flemish as L1 and had been brought up bilingual in Lebanon.

Table 5.31. Distribution of parents’ first language.

<table>
<thead>
<tr>
<th></th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic as L1</td>
<td>82</td>
<td>73</td>
</tr>
<tr>
<td>Other language as L1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Bilingual Arabic and French or English</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Single parent household (no Parent2 data)</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Missing information</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Similarly to the cross-sectional study in Sweden, the large majority of the parents in Lebanon had also been born in an Arabic-speaking country (Lebanon) and had Arabic as their L1.

5.3.4 Language use in the home

The parental questionnaire also targeted information regarding language use between various family members in the home: language spoken between the parents, the parents’ language to the child, the child’s language to the parents and the child’s language to the siblings.

The language the parents speak with each other at home gives an idea of what language the child is exposed to (even if not directly addressed) on a daily basis. The majority of the parents reported that they spoke almost only Arabic at home.\footnote{Code-switching of everyday terms and expressions is probably not taken into account (for example, the common use of the word \textit{computer} instead of \textit{ḥasūb} in Arabic, or \textit{ascenseur}, ‘elevator’ in French, instead of \textit{maṣāḥ} in Arabic, or the casual greeting phrase \textit{hi, kifak, qa va? ‘hi (in English), how are you (in Lebanese Arabic), is it going well (in French)?’}.} Thirteen parents reported that they spoke a combination of Arabic and a second language (English and/or French) with each other. Only
one family (parents to BiLeb6-23) indicated that the parents spoke solely French with each other at home. Table 5.32 summarizes the parents’ language use between each other at home.

**Table 5.32.** Parents’ language use with each other.

<table>
<thead>
<tr>
<th>Only Arabic</th>
<th>Arabic and English /French</th>
<th>Only English /French</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>13</td>
<td>1</td>
<td>14</td>
<td>100</td>
</tr>
</tbody>
</table>

The language that the parents speak with the child on a daily basis at home provides a direct input source to the child. Table 5.33 shows the parents’ reported language use with their child. Unlike the cross-sectional study in Sweden where the majority of the parents reported that both parents spoke mainly Arabic at home \( (N = 78) \), only 35% of the parents in Lebanon reported that they spoke mainly Arabic at home with the child. Twenty-two households (22%) reported that the parents spoke both languages in roughly equal amounts with their children.

**Table 5.33.** Parents’ language with the child.

<table>
<thead>
<tr>
<th>Both parents mainly Arabic</th>
<th>Both parents mainly English/French</th>
<th>1 parent mainly Arabic, 1 parent 50/50</th>
<th>1 parent mainly Arabic, 1 parent mainly English/French, 1 parent mainly 50/50</th>
<th>Both parents 50/50</th>
<th>Other</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>11</td>
<td>12</td>
<td>8</td>
<td>22</td>
<td>0</td>
<td>11</td>
<td>100</td>
</tr>
</tbody>
</table>

*Note. 50/50 = ca. 50% Arabic, ca. 50% English/French.*

The parents were also asked to report which language(s) the child speaks to them. Table 5.34 shows the children’s reported language use with their parents. In general, the children’s reported language use with the parents mirrored that of the parents’ language use with them.

**Table 5.34.** Children’s language use with parents.

<table>
<thead>
<tr>
<th>To both parents mainly Arabic</th>
<th>To both parents mainly English/French</th>
<th>To 1 parent mainly Arabic, with 1 parent 50/50</th>
<th>To 1 parent mainly Arabic, with 1 parent mainly English/French</th>
<th>To 1 parent mainly English/French, with 1 parent 50/50</th>
<th>To both parents 50/50</th>
<th>Other</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>18</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>0</td>
<td>11</td>
<td>100</td>
</tr>
</tbody>
</table>

*Note. 50/50 = ca. 50% Arabic, ca. 50% English/French.*
Recall that in the cross-sectional study in Sweden, the children’s language use to their parents also often mirrored that of their parents. However, the parental language use to the child in Sweden and Lebanon do not fully resemble each other, hence, children’s language use in the two cross-sectional studies does not either. More parents and children in Lebanon were reported to speak with each other a combination of Arabic and a second language (50/50) than what was reported in Sweden (compare Tables 5.33 and 5.34 with Tables 5.9 and 5.10, respectively). Furthermore, more children in Lebanon ($N = 18$) were reported to speak mainly the second language to both their parents, compared to children in Sweden ($N = 8$).

As for the children’s language use with sibling(s), forty-one children were reported to speak both Arabic and a second language to their sibling, as can be seen in Table 5.35. Hence, the children’s language use with the sibling(s) in Lebanon does not mirror language use with the parents, where far fewer children ($N = 19$) were reported to speak both Arabic and a second language.

<table>
<thead>
<tr>
<th>No sibling</th>
<th>Mostly Arabic</th>
<th>Mostly English/French</th>
<th>Other English/French</th>
<th>Missing information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>21</td>
<td>41</td>
<td>17</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

The children’s language use with their sibling(s) does, however, resemble that of the language use of the Arabic-speaking bilingual children in Sweden, who also spoke a combination of two languages with the siblings (recall Table 5.11). However, fewer children in Lebanon were reported to speak mostly Arabic with their siblings in comparison to the children in Sweden. In Lebanon, twenty-one children were reported to speak mostly Arabic with their siblings, whereas almost twice as many children ($N = 39$) spoke mostly Arabic with their siblings in Sweden.

Parents were also asked to specify their child’s birth order. The parents chose the number that indicated the order in which their child was born with respect to the siblings. As can be seen in Table 5.36, a little over half the children ($N = 52$) are first-borns (do not have an older sibling), whereas 39 children have at least one older sibling. For 9 children, information regarding child’s birth order was unavailable.

<table>
<thead>
<tr>
<th>1st born</th>
<th>2nd born</th>
<th>3rd born</th>
<th>4th born</th>
<th>5th born</th>
<th>6th born</th>
<th>7th born</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>32</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

Recall that several of the children participating in the Swedish cross-sectional study were also first born amongst their siblings ($N = 39$), however, the majority had at least one older sibling ($N = 61$). In the Lebanese sample, there
were no children who were 4\textsuperscript{th}, 5\textsuperscript{th}, 6\textsuperscript{th} or 7\textsuperscript{th} born, in contrast to the Swedish sample where there was a total of 12 children whose birth order was the 4\textsuperscript{th} or up to the 7\textsuperscript{th} born child.

On a 4-point scale, parents were asked to specify how often they read books with their child at home: ‘almost everyday’, ‘once/twice a week’, ‘twice a month’ or ‘(almost) never’. In Table 5.37, 44 households reported that shared book reading activities occurred at least once a week in Arabic, whereas 66 reported that they read in English or French. Note, however, the large amount of missing information on this question.

**Table 5.37.** Parent and child joint book reading activity at home in Arabic and a second language.

<table>
<thead>
<tr>
<th></th>
<th>Almost everyday</th>
<th>Once/twice a week</th>
<th>Twice a month</th>
<th>Almost never</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading in Arabic</td>
<td>20</td>
<td>24</td>
<td>8</td>
<td>15</td>
<td>33</td>
</tr>
<tr>
<td>Reading in English/French</td>
<td>34</td>
<td>32</td>
<td>6</td>
<td>6</td>
<td>22</td>
</tr>
</tbody>
</table>

In both the Swedish and the Lebanese cross-sectional samples, there were high proportions of missing data for this question regarding joint book reading at home, but more so in Lebanon than in Sweden. One possible explanation as to why there was a large number of missing information on this particular question could be related to the fact that parents do not want to ‘reveal’ that they do not do this activity at home.

Interestingly, more families in Sweden reported that they read frequently with their child in Arabic in comparison to the number of families in Lebanon who instead reported that they read in the second language more than they read in Arabic.

### 5.3.5 Estimated daily language exposure

A little less than one third of the children ($N = 30$) were reported to be exposed to an even amount (50/50) of the languages throughout the day. About one third of the children ($N = 32$) were reported to have the majority of their language exposure in Arabic, and the remaining children ($N = 27$) were reported to have the majority of their daily language exposure to English/French. No child was reported to have any ‘other’ estimated language exposure. Table 5.38 reports the estimated daily language exposure of the children in the cross-sectional study in Lebanon according to the parents. The distribution of daily language exposure in the Lebanese sample is similar to that in the Swedish cross-sectional sample except for one aspect.
Table 5.38. Estimated daily language exposure as per age group.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Eng./Fre. 5% Arabic</th>
<th>Eng./Fre. 20% Arabic</th>
<th>Eng./Fre. 40% Arabic</th>
<th>Eng./Fre. 50% Arabic</th>
<th>Eng./Fre. 60% Arabic</th>
<th>Eng./Fre. 80% Arabic</th>
<th>Eng./Fre. 95% Arabic</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>9</td>
<td>21</td>
<td>30</td>
<td>12</td>
<td>15</td>
<td>0</td>
<td>11</td>
</tr>
</tbody>
</table>


Twice as many children in Lebanon (N = 15) were reported to have 80% exposure to English/French and only 20% exposure to Arabic. In contrast, only seven children in the Swedish cross-sectional study were reported to have the majority of their daily language input in Swedish (Swedish 80%, Arabic 20%).

5.3.6 Attitudes towards language

The parents were asked to report their children’s language preference and which language the parents considered was the most important for the child to learn.

Regarding the children’s language preference, about one third of the children (N = 36) did not prefer one language over the other, and another third (N = 35) were reported to prefer speaking a second language over Arabic. A relative small number of children (N = 12) were reported to prefer speaking Arabic. Two five-year-old children (the twins BiLeb5-16 and BiLeb5-17) were reported to prefer speaking German over any other language. Table 5.39 provides an overview of the children’s language preference. The answers were evenly distributed across all age groups.

Table 5.39. Child’s language preference.

<table>
<thead>
<tr>
<th>None</th>
<th>Arabic</th>
<th>English/French</th>
<th>Other</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>12</td>
<td>35</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>

When parents were asked about their attitude towards which language was the most important for their child to learn, they were given the option to choose between ‘Arabic’, ‘English’, ‘French’, ‘all of them’ and ‘other’. The majority of the parents (N = 59) reported that all the languages (Arabic, English and French) were equally important for their children to learn. Few parents (N = 9) considered Arabic to be the most important language, likewise, few parents (N = 15) stated that learning English and/or French (without the necessity of learning Arabic) was the most important. In the ‘other’ category, no parent
indicated a new language, but rather another language combination. Two parents considered learning ‘Arabic and French’ to be the most important, whereas four parents ranked ‘Arabic and English’ highest.

Table 5.40. Parents valuing their child’s language(s).

<table>
<thead>
<tr>
<th>All (Arabic and English and French)</th>
<th>Arabic</th>
<th>English and/or French</th>
<th>Other</th>
<th>Missing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>9</td>
<td>15</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

5.3.7 Summary of the participants in the cross-sectional study in Lebanon

Information about the children who participated in the cross-sectional study in Lebanon ($N = 100$, age 4;0–8;1) was gathered via the same parental questionnaire as in Sweden. Notably, in Lebanon, a relatively high number of questionnaires were missing. Background information about the children, their parents, language use within and outside the family, and the child’s language exposure patterns are summarized below.

The majority of the children ($N = 87$) were born in Lebanon and 4 children were born in a non-Arabic speaking country. The children spoke Lebanese Arabic except for two children who also spoke Syrian and Palestinian, respectively. Unlike in the cross-sectional study in Sweden, not all children had started to receive regular exposure to Arabic at birth, however the majority of them had ($N = 72$). Age of onset to a second language (English or French) occurred quite early (before the age of 3) for most children ($N = 77$). Thirty-nine children had at least one older sibling. (In the cross-sectional sample in Sweden, more than half of the children ($N = 61$) had an older sibling.)

The majority of the parents were born and raised in Lebanon. Few parents had been born outside of Lebanon but brought up in Lebanon, another few had been born in Lebanon but brought up elsewhere, and two parents were born and raised in an Arabic-speaking country. The parents’ first language was Arabic (except for 2 parents whose L1 was another language, and 1 parent who was raised bilingual speaking Arabic and English or French). The parents’ SES (education) ranged between lower secondary education and doctoral degrees. The mean level of education was equivalent to having started or finished a university BA degree, which is higher than the mean education of the parents in the Swedish cross-sectional study. A large amount of data (about one fourth of the parents) related to this question was missing in the Lebanese cross-sectional study.

138 In contrast to the cross-sectional study in Sweden, calling the parents to complement missing questions or asking for a phone interview was not possible. As previously mentioned, data collection was limited to two short periods in time. Additional extended contact with the parents after that time was not possible.
Regarding languages spoken at home, the majority of the parents \((N = 72)\) indicated that they spoke only Arabic with each other while 13 couples said that they spoke both Arabic and English/French. One couple spoke only the second language at home (and no Arabic). This language use did not mirror the parent’s language choice with the child where only about one third of the parents \((N = 35)\) indicated that both parents spoke mainly Arabic with the child while 11 parents indicated that they spoke mainly English/French with the child. This does not match the language use of the parents in Sweden where more parents spoke Arabic with each other and the majority of the parents spoke Arabic with the child. Furthermore, no couple chose to speak predominantly the second language (Swedish) at home in Sweden. In Lebanon, 22 households reported that the parents spoke an even amount of Arabic and English/French with their child. By comparison in Sweden, only 9 couples spoke an even amount of Arabic and Swedish with their child. The language that the children in Lebanon spoke with their parents generally mirrored that which the parents spoke with them.

As for language use with siblings, 41 children were reported to speak both Arabic and English/French, while 17 spoke English/French, and only 21 spoke mainly Arabic. The children’s language use with siblings was quite similar in the cross-sectional studies in Sweden and Lebanon. As for joint book reading, less than half of the households \((N = 44)\) in Lebanon reported that the parents read with their child (in Arabic) at least once a week. In the Swedish cross-sectional study, the number of parents who read in Arabic with their children was a little higher \((N = 53)\). For estimated daily language exposure, about one third of the children were reported to receive an even amount of exposure to Arabic and English/French throughout the day, while one third were exposed mostly to Arabic and another one third were exposed mostly to the second language (English/French).

Finally, regarding the children’s language preference, only 36 children were reported not to have any preference, and almost as many children \((N = 35)\) were reported to prefer the second language (English/French) over Arabic. Only 12 children were indicated to prefer Arabic. This does not resemble the language preference pattern of the children in the Swedish cross-sectional study, where about half of the children were reported to have no language preference and about one fourth of the children preferred Arabic and another one fourth preferred the second language (Swedish). While the vast majority of the parents in Sweden \((N = 81)\) indicated that they preferred that their child learnt both Arabic and Swedish, and only two parents preferred Swedish over Arabic, the majority of the parents in Lebanon \((N = 59)\) reported that they prefer their child to learn three languages: Arabic, English and French, indicating the importance of all three. Fifteen parents stressed the importance of English and/or French over Arabic, while only 9 parents had a preference for Arabic.
In the present chapter (Chapter 5), an extensive overview of the participants and their families has been presented for the cross-sectional and longitudinal studies in Sweden and the cross-sectional study in Lebanon. In the coming chapters, the results of the children will be presented in relation to their performance on the lexical and narrative macrostructure tasks, starting with the cross-sectional study in Sweden and vocabulary.
6. Cross-sectional study in Sweden: Vocabulary

This chapter reports the results for vocabulary comprehension and vocabulary production for the cross-sectional study in Sweden and is divided into four main sections. Section 6.1 contains the research questions. Operationalization of variables and statistical analyses (6.2) are described thereafter. Section 6.3 reports the results starting with an overview (6.3.1), followed by the results for vocabulary and age (6.3.2) and vocabulary and background factors (6.3.3). The chapter ends with a discussion (6.4).

6.1 Research questions
6.1.1 Vocabulary scores and age
The following research questions concerning vocabulary and age are investigated:
- Is there a difference between the vocabulary scores in Arabic and Swedish?
- How does vocabulary develop with age in Arabic and Swedish?
- Is there a difference between the comprehension scores and the production scores in each language?

6.1.2 Vocabulary scores and background factors
The following research questions concerning vocabulary scores and background factors are explored:
- Is there a relationship between the age of onset of Swedish and Arabic and Swedish vocabulary comprehension and production?
- How does language use in the home affect Arabic and Swedish comprehension and production?
- How does language use outside the home affect Arabic comprehension and production?
- How do age and background factors combined affect Arabic and Swedish vocabulary?
6.2 Operationalization of variables and statistical analyses

6.2.1 Operationalization of variables

All background information about the participants (and their families) was gathered from the parental questionnaire (see Chapter 4 and Chapter 5). What follows is a description of how each of the background factors was operationalized.

For variables where information from two parents was needed (SES, parents’ language use with the child, child’s language use with the parents) but where only information from one parent was provided the value for the one parent was used. There were 5 such cases.

Age of onset (AoO) was obtained by asking the parents to indicate the age at which their child started to receive regular input in each of their languages, Arabic and Swedish. For Arabic, all children, except for one child, had regular input in Arabic from birth. (For one other child, no information was available, for neither language.) Because of this lack of variation, no statistical analyses on AoO for Arabic were run. For Swedish, the children’s age of onset varied greatly and they accordingly were grouped into 2 categories: children who had received regular input in Swedish before the age of 3 (0–2;11, \(N=48\)) and children who had received regular input in Swedish starting from the age of 3 or sometime later (3;0–7;0, \(N=51\)).

Socio-economic status (SES) was operationalized as parental education (see Section 5.1.3). In the questionnaire, each parent was asked in an open-ended question to report their highest level of education. All answers were then coded according to the International Standard Classification of Education (ISCED) 2011 on a 9-level scale (UNESCO Institute for Statistics, 2012). Later, for every child, an averaged score was created by combining each parent’s education score and dividing their sum by two. Information regarding parental education was missing from 17 parents. Children whose neither parent had filled in an answer for the education question (\(N=6\)) were excluded from the SES analyses.\(^{140}\) The children were classified into one of two groups (binary split): low-SES (ISCED 0–3, non-completed primary school up to completed secondary education, \(N=30\)) vs. high-SES (ISCED 4–8, post-

---

\(^{139}\) This child (BiAra5-25) had received regular input in Arabic from age one.

\(^{140}\) Mean score substitution (as proposed by Widaman (2006)) was also considered as a method to deal with SES missing data. This would have meant that the children whose both parents had left the education question empty (\(N=6\)) were given an SES value of 4.3 which was the mean SES of the whole sample (the mean of all the parents’ education values). Similarly, the two children who had missing information from one of their parents, were given the mean SES sample value for that parent. When analyses were made on all three types of SES data categorization (binary split, three-way split, linear), no significant differences were observed between the analyses with mean score substitution and those without. The statistics with mean score substitution of missing data will not be reported here. Instead, see Section 6.3.3.2 for the results of the SES analyses where children with missing SES data were excluded.
secondary non-tertiary education up to completed doctorate, \( N = 64 \). The cut-off was placed so that the value 3.5 belonged to the high-SES group.

As for the parents’ language use to the child, parents were asked to estimate on a five-point scale how much Arabic and Swedish each parent spoke to the child.\(^{141}\) The scale ranged from ‘(almost) only Arabic’, ‘mostly Arabic, sometimes Swedish’, ‘Arabic 50% – Swedish 50%’, ‘mostly Swedish, sometimes Arabic’ to ‘(almost) only Swedish’. Parents could also indicate an ‘other’ category where they specified whether they spoke other languages or language combinations to the child. All children were then classified into two categories: ‘only/mostly Arabic’ \( (N = 79) \) if both parents had indicated that they spoke ‘(almost) only Arabic’ or ‘mostly Arabic, sometimes Swedish’ to the child, vs. ‘other’ \( (N = 20) \) if parents used any other language combination at home (for example, if one parent spoke ‘(almost) only Arabic’ and the other spoke ‘Arabic 50% – Swedish 50%’).

Regarding the child’s language use with the parents, the same five-point-scale and categorical division were used. If a child spoke ‘only/mostly Arabic’ \( (N = 68) \) to both their parents, s/he was classified into one group, and if s/he spoke another language combination \( (N = 31) \), s/he was classified into the ‘other’ group.\(^{142}\)

As for the child’s language use with sibling(s), parents were asked to indicate on a three-point-scale whether their child spoke ‘mostly Arabic’, ‘Arabic 50% – Swedish 50%’ or ‘mostly Swedish’ with their sibling(s). Parents could also write down another language combination in an ‘other’ category, or specify that their child had no sibling. Six children had no siblings. These children were excluded from the analyses. Information was missing from 1 child. The children were categorized in a binary fashion allowing for two comparative analyses: In the first analysis, children who spoke ‘mostly Arabic’ \( (N = 39) \) were compared with children who spoke ‘other’ language combinations with their sibling(s) \( (N = 54) \). In the second analysis, children who spoke ‘mostly Swedish’ \( (N = 18) \) were compared with children who spoke ‘other’ language combinations with their sibling(s) \( (N = 75) \).

An additional analysis was made to investigate whether being the oldest sibling had any significant effect on vocabulary compared to having (an) older sibling(s). Parents were asked to specify birth order. Children were classified into two groups: children who were first-born (i.e. had no older sibling(s), \( N = 39 \)) and those who had at least one older sibling \( (N = 61) \). Children who did not have any siblings \( (N = 6) \) were not included in the analysis since it was the child’s possibility to speak to older or younger siblings that was investigated rather than being an only child.

---

\(^{141}\) Information about language use from parents to child and from child to parents was missing for 1 child.

\(^{142}\) No extra split of ‘only/mostly Swedish’ was made since only very few children \( (N = 8) \) belonged to this group.
Parents were asked whether they read any books in Arabic and/or Swedish with their child. Parents could choose one of four options: reading ‘almost everyday’, ‘once/twice a week’, ‘twice a month’ and ‘(almost) never’. For each language, the answers were categorized into two groups: ‘reading in X often’ (‘almost every day’ or ‘once/twice a week’) or ‘reading in X rarely’ (‘twice a month’ or ‘never’) where X refers to either Arabic or Swedish. For the Arabic joint book reading activity, a little more than half of the parents (N = 53) were reported to read Arabic stories ‘often’ with their child, and one fourth of the parents (N = 25) ‘rarely’ performed any Arabic reading activities with their child at home. For Swedish, about half of the parents (N = 49) reported that they read ‘often’ in Swedish with their child, and a little more than one third (N = 34) reported that they ‘rarely’ did.\textsuperscript{143}

Regarding the child’s language input from extended family members, friends/playmates and staff members at school, parents were asked to answer ‘yes-or-no’ questions whether their child heard the home language (Arabic) from such persons outside the home. For each question, the children were then classified into ‘yes’ or ‘no’ groups accordingly: children who heard Arabic from extended family members (N = 78) vs. children who did not (N = 21); children who heard Arabic from friends or playmates (N = 29) vs. children who did not (N = 70); and children who heard Arabic from school staff members (N = 25) vs. children who did not (N = 72).\textsuperscript{144}

Parents were also asked whether their child attended any form of mother tongue instruction (MTI) (see Section 5.1.7). Based on the parents’ answers,\textsuperscript{145} children were assigned to two categories: ‘yes, attends MTI’ (N = 65) or ‘no, does not attend MTI’ (N = 31).

Finally, parents were asked to estimate their child’s daily language exposure to each language on a seven-point-scale ranging from ‘95% Arabic – 5% Swedish’, ‘80% Arabic – 20% Swedish’, ‘60% Arabic – 40% Swedish’, ‘50% Arabic – 50% Swedish’, ‘40% Arabic – 60% Swedish’, ‘20% Arabic – 80% Swedish’ to ‘5% Arabic, 95% Swedish’. The children were then categorized into three groups: ‘mostly Arabic’ (60%–90% Arabic, N = 30), ‘even exposure’ (50% Arabic – 50% Swedish, N = 37) and ‘mostly Swedish’ (60%–90% Swedish, N = 30). Two children were excluded from the analyses since their parents had filled in the ‘other’ category, indicating that their child’s language

\textsuperscript{143} Regarding joint parent-and-child book reading activities at home, for Arabic reading and Swedish reading, information was missing for 22 and 17 children, respectively.

\textsuperscript{144} For home language use with extended family and friends, there was missing information for 1 child. For whether the child heard Arabic from staff members, no information was available for 3 children.

\textsuperscript{145} Some parents indicated that their children attended MTI lessons provided by the municipality, others via private initiative, and others indicated that their children attended both (recall Section 5.1.7). All of these options were grouped into the ‘yes’ category.
exposure patterns did not match any of the groups above.\textsuperscript{146} The daily language exposure variable was re-categorized for the regression analyses to become a binary variable by grouping two of the variables into one under the category of ‘other’. For example, when wanting to analyse the effect of ‘mostly Swed-
dish’ language exposure, ‘even’ and ‘mostly Arabic’ language exposure vari-
ables were grouped into one to form an ‘other’ category.

6.2.2 Statistical analyses
The level of significance was set constant at $p < .05$ (two-tailed) for all statis-
tical analyses.

6.2.2.1 Vocabulary scores and age
The same analyses were performed on both Arabic and Swedish vocabulary scores, separately. The relationship between comprehension and production scores for each language was analysed with Pearson correlation (two-tailed). For each age group, paired samples t-tests were used to compare Arabic and Swedish comprehension scores and production scores.

Age effects for comprehension and production scores were calculated using one-way ANOVAs for the age groups. Post hoc analyses (with Bonferroni correction) were then used to identify which age groups differed significantly from each other.

When age was operationalized as a continuous variable (in months), simple regression analyses were run. The regression analyses were also visualized as regression lines on scatterplots that showed the children’s individual scores in relation to age.

6.2.2.2 Vocabulary scores and background factors
First, background factors were analysed to investigate their individual effect on vocabulary scores in each language. The following factors were analysed for both Arabic and Swedish separately: age of onset (AoO) of Swedish, socio-economic status (SES), parents’ language use to the child, child’s language use with the parents, child’s language use with sibling(s), joint book reading and estimated daily language exposure. Factors that were only ana-
lysed for Arabic vocabulary scores were: child’s home language use with ex-
tended family members, with friends/playmates, and with school staff mem-
bers, as well as mother tongue instruction (MTI) attendance.

Vocabulary scores in relation to background factors with a binary (two-
way) split were analysed using Welch’s independent samples t-tests (Welch’s,

\textsuperscript{146} One child (BiAra7-11) was reported to have had daily input (to some extent) in Kurdish (Sorani) from one parent. Another child (BiAra7-17) was reported to also have input in English at home.
equal variance not assumed). The binary background factors are AoO (of Swedish), SES (2-way split)\textsuperscript{147}, joint book reading, parents’ language to the child, child’s language with parents, with siblings, with extended family members, with friends/playmates and with school staff members. Estimated daily language exposure was analysed with a three-way split using one-way ANOVAs. Post hoc analyses (with Bonferroni correction) were then used to identify significant differences between the groups.

Second, multifactorial analyses in the form of linear regression analyses were conducted on each of the four vocabulary measures (Arabic comprehension, Arabic production, Swedish comprehension and Swedish production) in order to examine the combined effects of various background factors on the vocabulary measures. Each of the four linear regression models included only input factors that had individually proved to have a significant effect on the vocabulary measure in question. For example, when Arabic comprehension was the dependent variable, the independent variables used in the linear regression analysis were only those factors that proved to have a significant effect on Arabic comprehension when analysed individually. See Section 6.3.3.6 for additional information on which factors were excluded from the regression analyses.

In each of the four regression models, the following is reported for all independent variables: unstandardized coefficient ($B$), standard error ($SE$), standardized coefficient ($\beta$), and the p-value ($p$). The adjusted R-squared (adjusted $R^2$), the F-statistic (F) and its p-value are also reported in the models. The adjusted R-squared measures the strength of the relationship between the regression model and the dependent variables (i.e. the percentage of the variance in the dependent variable that the independent variables explain collectively).

6.3 Results

6.3.1 Overview of the Arabic and Swedish vocabulary results

This section provides an overview of the comprehension and production scores for Arabic and Swedish vocabulary. Table 6.1 displays the mean, standard deviation (SD) and ranges for both Arabic and Swedish for all children.

\textsuperscript{147} The effect of SES (parental education) was also explored in other ways. When SES was investigated as a continuous variable, a Pearson correlation (two-tailed) was run to analyse the relationship between vocabulary scores and SES. When SES was split into three-ways (high-mid-low), ANOVAs were used.
Table 6.1. All ages combined (4;0–7;11, N = 100) vocabulary (CLT) scores in Arabic and Swedish.

<table>
<thead>
<tr>
<th></th>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>45.6</td>
</tr>
<tr>
<td>SD</td>
<td>7.5</td>
<td>10.8</td>
</tr>
<tr>
<td>Range</td>
<td>25–59</td>
<td>18–60</td>
</tr>
<tr>
<td>Production scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32.7</td>
<td>30.8</td>
</tr>
<tr>
<td>SD</td>
<td>12.3</td>
<td>11.7</td>
</tr>
<tr>
<td>Range</td>
<td>1–53</td>
<td>10–53</td>
</tr>
</tbody>
</table>

*Note.* Max comprehension score = 60 points, max production score = 60 points.

Arabic and Swedish vocabulary comprehension and production scores were compared with paired samples t-tests. At group level, no statistically significant difference was found between the comprehension scores ($t(99) = 1.71, p = .090$) and between the production scores ($t(99) = 1.04, p = .301$). Unsurprisingly, in both languages, the comprehension scores were higher than the production scores (Arabic: $t(99) = 18.51, p < .001, d = 1.85$, large effect size; Swedish: $t(99) = 30.56, p < .001, d = 3.06$, large effect size). For both Arabic and Swedish, the SDs were smaller for comprehension than for production, thus indicating a more varied performance in production than in comprehension scores.

For comprehension, Swedish scores have both a wider range and a larger SD than the Arabic scores. For production, it is the other way around, Arabic scores have both a wider range and a larger SD than Swedish scores. In other words, in comprehension, the children (as a group) scored more similarly to each other in Arabic than in Swedish, whilst for production, it was the opposite, the children scored more similarly to each other in Swedish than in Arabic.

In order to see whether the order of testing had any effect on the children’s vocabulary scores, independent samples t-tests were run. There was no significant effect for order of testing for any vocabulary measure (Arabic comprehension: $t(97.53) = .750, p = .445$; Arabic production: $t(98.00) = 1.128, p = .262$; Swedish comprehension: $t(96.90) = -.910, p = .365$; Swedish production: $t(98.00) = -1.333, p = .186$). In other words, children performed equally well irrespectively of whether they were first tested in Arabic or whether they

---

148 See Figure 6.2 for a visualisation of the Arabic production scores in the form of a scatterplot. One 4-year-old child (BiAra4-02) scored particularly low (score = 1) in Arabic production, thus extending range.

149 Independent samples t-tests were also run on each age group separately to see whether order of testing affected a particular age group more than another. No significant differences were found irrespective of which language the children were tested in first. The statistics for each age group are not reported here.
were first tested in Swedish. Scores of individual items (target words) in Arabic production and Swedish production varied to some extent; however, item effects will not be discussed in the present study.

6.3.2 Vocabulary scores and age
This section is divided into four sections. First, vocabulary (CLT) comprehension and production scores are presented in relation to age for Arabic (Section 6.3.2.1) and for Swedish (Section 6.3.2.2). At the end of each language section, examples of individual children whose scores considerably deviated from their peers are discussed. Then, a comparison is made between the Arabic and Swedish comprehension and production scores (Section 6.3.2.3). The section ends with a brief summary of the main results (Section 6.3.2.4).

6.3.2.1 Arabic vocabulary scores and age
This section reports the Arabic vocabulary comprehension and production scores in relation to age.

Table 6.2. Arabic vocabulary (CLT) scores for each age group.

<table>
<thead>
<tr>
<th></th>
<th>4-year-olds</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>7-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp. N</td>
<td>22</td>
<td>25</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Mean</td>
<td>41.6</td>
<td>46.7</td>
<td>48.5</td>
<td>52.4</td>
</tr>
<tr>
<td>SD</td>
<td>7.6</td>
<td>6.5</td>
<td>7.5</td>
<td>3.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>4-year-olds</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>7-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production</td>
<td>Mean</td>
<td>25.6</td>
<td>32.6</td>
<td>34.5</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>12.2</td>
<td>11.7</td>
<td>13.2</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>1–42</td>
<td>11–48</td>
<td>10–53</td>
</tr>
</tbody>
</table>

Note. Comp. = comprehension. Max comprehension score = 60 points, max production score = 60 points.

As for whether the location of the testing session (at home, at (pre)school or at an organization/congregation) and the language used there could have contributed to the children’s performance, it was not possible to classify the testing locations into ‘predominate use of Arabic’ vs. ‘predominate use of Swedish’ since several of the testing locations were too diverse in terms of language use. For example, at some (pre)schools, children had the opportunity to hear much more Arabic than what other children heard at their preschools. The same goes for the congregations and associations where Arabic and Swedish use varied greatly from one location to the other. Hence, no statistical analysis will be reported regarding the effect of the testing location. Some test items showed a ceiling effect at all ages (for example, dricka ‘to drink’ in Swedish; sāʿa ‘a clock’ in Arabic) while other test items showed a floor effect (for example, posta ‘to post’ in Swedish and mizān ‘a scale’ in Arabic). Furthermore, despite the continuous efforts to establish a monolingual testing session, certain test items were more often code-switched that other test items (for example, gunga ‘to swing’ in Swedish instead of yitmarğah in Arabic; būme ‘owl’ in Arabic instead of uggla in Swedish). Item effects and different types of wrong answers are not investigated in the present study, but see Section 9.3.2.3 for a brief overview of code-switching in the Swedish and Lebanese cross-sectional samples.
Age development is investigated both for age groups (one-way ANOVAs) and linearly (regression analyses and scatterplots). Table 6.2 shows an increase in mean scores for both Arabic comprehension and production across the ages. One-way ANOVA tests confirm this pattern and show a significant difference in the scores between the age groups (Arabic comprehension: \( F(3,96) = 10.932, p < .001, \eta^2 = .255, \) large effect size; Arabic production: \( F(3,96) = 4.075, p = .009, \eta^2 = .113, \) medium effect size). The eta-squared (\( \eta^2 \)) values suggest that the effect of age is stronger for comprehension than for production in Arabic.

For the comprehension scores, post hoc analyses (with Bonferroni correction) revealed a significant difference between the 4-year-olds and all the other age groups (4-year-olds vs. 5-year-olds: \( p = .049, 4\)-year-olds vs. 6-year-olds: \( p = .002, 4\)-year-olds vs. 7-year-olds: \( p < .001 \)). There was only one additional significant difference between the remaining age groups, namely the 5-year-olds vs. 7-year-olds (\( p = .018 \)).

For the production scores, post hoc analyses (with Bonferroni correction) revealed that no significant difference existed between any age groups except for the 4-year-olds vs. 7-year-olds (\( p = .007 \)).

For both comprehension and production, the lowest score was found in the youngest age group (4-year-olds) and the highest score for comprehension was found in the 7-year-old group and for production in the 6-year-old group. Similarly, for both comprehension and production scores, the oldest group (7-year-olds) had the smallest SD and the narrowest range compared to the younger age groups.

Next, the relationship between vocabulary scores and age as a continuous variable is investigated. A simple linear regression analysis was carried out to examine whether age (in months) significantly predicted Arabic comprehension scores. The results of the regression analysis indicated significance (\( F(1,98) = 32.247, p < .001 \)) with an R-squared value of \( R^2 = .248, \) meaning that 24.8% of the variation of scores can be explained by the child’s age (in months). Figure 6.1 illustrates age development in form of a regression line drawn on a scatterplot of Arabic comprehension (CLT) scores against age in months.

---

152 Post hoc analyses (with Bonferroni correction) showed, however, that the difference between the 4-year-olds vs. 6-year-olds nearly reached significance (\( p = .051 \)).
Figure 6.1. Scatterplot of Arabic vocabulary (CLT) comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

Four children, three 4-year-olds and one 5-year-old, scored below 50% (30 points) on Arabic comprehension (CLT). The majority of the children ($N = 80$) scored between 50%–90% correct (30–54 points). The remaining children ($N = 16$) scored above 90% (55–60 points) on Arabic comprehension (CLT). These latter children consisted of two 5-year-olds, seven 6-year-olds and seven 7-year-olds.

Next, the relation between Arabic production scores for the individual children and age (as a continuous variable) is analysed. A simple linear regression proved a significant relationship between Arabic production scores and age ($F(1,98) = 11.651, p = .001$) with an $R$-squared value of $R^2 = .106$, meaning that 10.6% of the variation of scores can be explained by the child’s age (in months). Figure 6.2 illustrates this age development in the form of a regression line on a scatterplot of Arabic production (CLT) scores against age in months. Observe that the $R$-squared value for Arabic production ($R^2 = .106$) is smaller than that for Arabic comprehension ($R^2 = .248$), meaning that age has a larger explanatory power for comprehension.
In contrast to the Arabic comprehension scores, which only had 4 children scoring below 50% (30 points), Arabic production scores had more than one third of the children (\(N = 37\)) scoring below 50%. These children were from all age groups and not just from the youngest group. All the remaining children (\(N = 63\)) scored between 50%–90%. No child scored at 90% or higher in Arabic production.

When investigating the relationship between comprehension and production scores in Arabic vocabulary (CLT), Pearson correlation (two-tailed) showed that there was a strong positive correlation between the comprehension scores and production scores (\(r = .779, N = 100, p < .001\)). Figure 6.3 plots the comprehension and production scores against each other for the individual children.

As can be seen in Figure 6.3, close to half of the children (\(N = 44\)) are found in the upper right quadrant. This means that these children have scored higher than 50% in both Arabic (CLT) comprehension and production. Fourteen (two 5-year-olds, seven 6-year-olds and seven 7-year-olds) of these children scored 75% (45 points) or higher in both Arabic comprehension and production. In the opposite quadrant (lower, left), four children are found, indicating that they scored low (below 50%) in both comprehension and production.

Below individual children who performed much lower than their age group peers in Arabic comprehension and/or Arabic production are discussed in relation to certain background factors that might explain their low performance.
Figure 6.3. Scatterplot of Arabic vocabulary (CLT) comprehension and production scores. The grey lines indicate a 50% mark (30 points) of the max score.

For Arabic comprehension, and in the 4-year-olds, three children (BiAra4-02, BiAra4-15, BiAra4-16) scored particularly low. BiAra4-02 had a hard time cooperating during the Arabic test session. (This child also scored low in Arabic production.) BiAra4-02 was also tested in the follow-up study where one of the child’s parents mentioned in the parental interview how the child, since a young age, preferred speaking Swedish over Arabic, and often spoke Swedish with all of his siblings. BiAra4-15 was reported to have the majority (60%) of her daily language exposure in Swedish. BiAra4-15 scored 9 points higher in Swedish than in Arabic in both vocabulary comprehension and production. Although BiAra4-16 was reported to have had the majority of his daily language exposure (60%) in Arabic, the child scored lower in Arabic than in Swedish in both comprehension and production. Furthermore, although not reported in the parental questionnaire, BiAra4-16 appears to have had frequent daily exposure to English in addition to Arabic and Swedish which was later revealed in the oral parental interview when the family participated in the longitudinal study. In the 5-year-old group, only BiAra5-06 scored lower than his peers did. This child was reported to have had an estimated 80% of daily language exposure to Swedish. As for the 6-year-olds, three children (BiAra6-10, BiAra6-11 and BiAra6-26) scored lower than their peers and were all reported to have had the majority of their daily language exposure (60% or more) in Swedish. While BiAra6-10 and BiAra6-11 scored above their age group mean in Swedish comprehension and Swedish production, BiAra6-26 did not. Observing the testing behaviour of BiAra6-26, the child had a hard time cooperating during the testing session. In the 7-year-
old group, no child stood out from the rest with remarkably low scores in Arabic comprehension.

For Arabic production, and in the 4-year-old group, six children scored considerably lower than their peers. BiAra4-02 (who also performed low in Arabic comprehension) scored only one point in Arabic production. This child, apart from not cooperating well during the Arabic testing session, was reported to prefer speaking Swedish at home. In the Arabic production part, about one third of BiAra4-02’s replies would have been scored as correct in the non-target language (Swedish). Children BiAra4-15 and BiAra4-16, who also scored low in Arabic comprehension, also scored remarkably low in Arabic production. BiAra4-15 often produced target words in Swedish instead, while BiAra4-16 often replied in English. The remaining 4-year-olds (BiAra4-09, BiAra4-14, BiAra4-21) were reported to have had the majority (60% or more) of their daily language exposure in Swedish. Also, BiAra4-21 seems to have had input in English since several of the child’s Arabic production replies were in English. As for the 5-year-olds, five children (BiAra5-01, BiAra5-06, BiAra5-12, BiAra5-15, BiAra5-16 and BiAra5-25) scored low in Arabic production and were reported to have had the majority (60% or more) of their daily language exposure in Swedish. As for the 6-year-old group, five of the lowest scoring 6-year-olds (BiAra6-03, BiAra6-08, BiAra6-09, BiAra10, BiAra6-11, BiAra6-19) were reported to have had the majority (60% or more) of their daily language exposure in Swedish. Finally, in the 7-year-old group, two children (BiAra7-12 and BiAra7-13) scored particularly low in comparison to their peers. BiAra7-12 was reported to have had almost all (95%) of her daily language exposure in Swedish, whereas BiAra7-13 was reported to have had an even (50%-50%) daily language distribution between her/his languages. BiAra7-13 was however reported to have had an unusually early AoO of Swedish (from birth) and was reported to often speak Swedish with her parents at home. Another four children (BiAra7-02, BiAra7-03, BiAra7-14, BiAra7-17) in the 7-year-old group scored relatively low, considerably below their age mean (M = 37.13) as well as below 50% (30 points) in Arabic production. Two of these children (BiAra7-03 and BiAra7-14) were reported to have had the majority (60%) of their daily language exposure in Swedish. BiAra7-02 was reported to have had an even daily language exposure, but scored just as low in Arabic production and just as high in the Swedish vocabulary measures as BiAra7-03 and BiAra7-14. As for BiAra7-17, the parents reported that the child had additional daily English exposure. The child used many English words in both of the Arabic and Swedish (CLT) production tasks. While none of BiAra7-17’s parents are native speakers of English, it is likely that the child has started to receive English exposure via television or internet videos.

In sum, for nearly all of the low-performing children in Arabic comprehension and production, it seems that an explanation may be found in their low exposure to Arabic. Statistical analyses of exposure effects on Arabic vocabulary scores are provided in Section 6.3.3.6.
6.3.2.2 Swedish vocabulary scores and age

This section reports the Swedish vocabulary comprehension and production scores in relation to age.

Table 6.3. Swedish vocabulary (CLT) scores for each age group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>4-year-olds</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>7-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>22</td>
<td>25</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Mean Comp.</td>
<td>36.1</td>
<td>44.9</td>
<td>46.9</td>
<td>53.3</td>
</tr>
<tr>
<td>SD</td>
<td>8.3</td>
<td>9.0</td>
<td>10.2</td>
<td>8.5</td>
</tr>
<tr>
<td>Range</td>
<td>18–52</td>
<td>29–59</td>
<td>27–60</td>
<td>27–60</td>
</tr>
<tr>
<td>Mean Prod.</td>
<td>22.3</td>
<td>29.1</td>
<td>31.6</td>
<td>39.5</td>
</tr>
<tr>
<td>SD</td>
<td>7.2</td>
<td>10.0</td>
<td>11.5</td>
<td>11.3</td>
</tr>
<tr>
<td>Range</td>
<td>10–41</td>
<td>15–48</td>
<td>11–48</td>
<td>12–53</td>
</tr>
</tbody>
</table>

Note. Comp. = Comprehension. Max comprehension score = 60 points, max production score = 60 points.

Age development is investigated both in age groups (one-way ANOVAs) and linearly (regression analyses and scatterplots). As can be seen in Table 6.3, there is an increase in mean scores for both Swedish comprehension and production across the ages. One-way ANOVA tests confirm this pattern and show a significant difference in the scores between the age groups (Swedish comprehension: $F(3,96) = 14.069, p < .001, \eta^2 = .305$, large effect size; Swedish production: $F(3,96) = 10.993, p < .001, \eta^2 = .256$, large effect size). The eta-squared ($\eta^2$) values suggest that the effect of age is somewhat stronger for comprehension than for production scores in Swedish. (Recall that for Arabic vocabulary, comprehension scores increased much more strongly with age than production scores.)

For the comprehension scores, post hoc analyses (with Bonferroni correction) revealed a significant difference between the 4-year-olds and all the other age groups (4-year-olds vs. 5-year olds: $p = .008$, 4-year-olds vs. 6-year-olds: $p < .001$, 4-year-olds vs. 7-year-olds: $p < .001$). An additional significant difference between the remaining age groups was found between the 5-year-olds vs. 7-year-olds ($p = .010$). These differences between the age groups for Swedish comprehension mirror those for Arabic comprehension. Recall Section 6.3.2.1. For production, post hoc analyses (with Bonferroni correction) revealed that the 4-year-olds scored significantly lower than the 6-year-olds ($p = .011$) and the 7-year-olds ($p < .001$) but not the 5-year-olds ($p = .154$). Also, there was a significant difference between the 7-year-olds and all the younger age groups (5-year-olds vs. 7-year-olds $p = .004$, 6-year-olds vs. 7-year-olds $p = .041$).

For comprehension, the lowest score was found in the youngest age group (4-year-olds) and the highest scores were found in the oldest age groups (6-year-olds and 7-year-olds). For production, the highest score was found
merely in the 7-year-old group. For both comprehension and production, the standard deviation (SD) was similar amongst all age groups, however, the 4-year-olds had a slightly smaller SD compared to the other age groups due to the lower top score.

In what follows, the relationship between Swedish vocabulary scores and age as a continuous variable is investigated. A simple linear regression analysis was carried out to test whether age (in months) significantly predicted Swedish comprehension scores. The results of the regression analysis proved significant \( F(1,98) = 35.059, p < .001 \) with an \( R\)-squared value of \( R^2 = .263 \), meaning that 26.3% of the variation of scores can be explained by the child’s age (in months). Figure 6.4 illustrates age development in the form of a regression line drawn on a scatterplot of Swedish comprehension (CLT) scores against age in months.

![Figure 6.4](image)

**Figure 6.4.** Scatterplot of Swedish vocabulary (CLT) comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

Spread across all age groups, a total of eleven children scored below 50% (30 points) on Swedish comprehension. The majority of the children \( (N = 63) \) scored between 50%–90% correct (30–54 points). The remaining children \( (N = 26) \) scored above 90% (55–60 points). These top performers consisted of four 5-year-olds, eight 6-year-olds and fourteen 7-year-olds.

As for the Swedish production scores, a simple linear regression analysis revealed a significant relation between Swedish production and age (in months) \( F(1,98) = 26.564, p < .001 \) with an \( R\)-squared value of \( R^2 = .213 \), meaning that 21.3% of the variation of scores can be explained by the child’s age. Figure 6.5 illustrates this age development in the form of a regression line on a scatterplot for Swedish production (CLT) scores against age in months.
The $R$-squared value for Swedish production ($R^2 = .213$) is somewhat smaller than that for Swedish comprehension ($R^2 = .263$), meaning that comprehension scores increase to some degree more strongly with age than production scores.

Almost half of the children ($N = 48$) scored below 50% on Swedish production (CLT). These children belonged to all age groups. The remaining children ($N = 52$) scored between 50%–90%. No child scored 90% or higher in Swedish production, which is very different from Swedish comprehension. Seven 7-year-olds scored quite high though (between 85%–88%).

Next, the relationship between comprehension and production scores is examined. A Pearson correlation (two-tailed) was used to investigate the relationship between comprehension and production scores in Swedish vocabulary (CLT). The results revealed a strong positive correlation between the comprehension scores and production scores ($r = .911$, $N = 100$, $p < .001$). Figure 6.6 visualizes the comprehension and production scores plotted against each other for the individual children.

As can be seen in Figure 6.6, about half of the children ($N = 52$) are found in the upper right quadrant, meaning that they scored higher than 50% in both Swedish comprehension and production. Thirteen (two 5-year-olds, four 6-year-olds and seven 7-year-olds) of these children scored even at 75% (45 points) or higher in both comprehension and production. In the opposite quadrant (lower, left), eleven children are found, indicating that they scored below 50% in both Swedish comprehension and production. Many of these children...
are 4-year-olds (N = 5), but children belonging to the older age groups can also be found (one 5-year-old, four 6-year-olds and one 7-year-old).

![Figure 6.6. Scatterplot of Swedish vocabulary (CLT) comprehension and production scores. The grey lines indicate a 50% mark (30 points) of the max score.](image)

Discussed below are individual children who performed considerably lower than their age group in Swedish comprehension and/or Swedish production in relation to certain background factors that might explain their low performance.

For Swedish comprehension, and in the 4-year-old group, one child (BiAra4-12) scored considerably lower than the rest of the children in his age group. This child was reported to have had the vast majority (80%) of his daily language input in Arabic. In the 5-year-old group, one child (BiAra5-05) scored below 50% (30 points). BiAra5-05 was also reported to have had an estimated 80% exposure to Arabic throughout the day. As for the 6-year-olds, four children (BiAra6-20, BiAra6-25, BiAra6-26, BiAra6-29) scored below 50% on the Swedish (CLT) production task. BiAra6-29 and BiAra6-25 were reported to have had an even daily exposure to Arabic and Swedish, whilst BiAra6-20 was reported to have had the majority (60%) of his daily language exposure in Arabic. These three children scored higher in their Arabic vocabulary tasks than in their Swedish. BiAra6-26, who was previously discussed in the Arabic vocabulary section, scored low on all vocabulary measures and was reported to have had a late AoO of Swedish (less than two years of Swedish exposure before the testing). Finally, in the 7-year-old group, two children (BiAra7-05 and BiAra7-16) scored considerably lower than their age
group. Both children were relative newcomers to Sweden and had not been exposed to Swedish for a long time.\textsuperscript{153}

For Swedish production, the scores were quite scattered in all age groups. In the 4-year-old group, one child (BiAra4-08) scored much lower than her peers. This child had received very high scores in Arabic comprehension and production, and was reported to have had the vast majority (80\%) of her daily language exposure in Arabic. Furthermore, BiAra4-08 was reported to have started to receive regular input to Swedish only one year prior to testing. In the 5-year-old group, two children (BiAra5-08 and BiAra5-20) scored considerably lower than their peers. BiAra5-08 and BiAra5-20 were reported to have had the majority of their daily language exposure in Arabic with 60\% and 95\%, respectively, and both children had had late onset of exposure to Swedish (after the age of 3). In the 6-year-old group, seven children scored considerably lower than their peers in Swedish production. Two of these children (BiAra6-06 and BiAra6-20) were reported to have the majority (60\%) of their daily language exposure in Arabic. Four of the remaining children (BiAra6-15, BiAra6-17, BiAra6-24, BiAra6-29) were reported to have an even daily language distribution. BiAra6-26, who has been previously mentioned, was reported to have had the majority (60\%) of his daily language exposure in Swedish, but had scored low in all four vocabulary measures.\textsuperscript{154} All the low-scoring 6-year-olds had an AoO of Swedish after the age of 3.\textsuperscript{155} Two of these children (BiAra6-25 and BiAra6-26) had only started to be exposed to Swedish at the age of 5 (i.e. approximately one year prior to testing). Finally, in the 7-year-old group, three children (BiAra7-05, BiAra7-16 and BiAra7-17) scored considerably lower than their peers in Swedish production. All three children had a late exposure to Swedish (after the age of 3).

As for the children’s daily language input, BiAra7-17 was reported to have had additional exposure to English, whilst BiAra7-16 had the majority (60\%) of her daily language exposure in Arabic. No information regarding language exposure was available for BiAra7-05.

Background factors that seem to be common for many of the low-scoring children in Swedish vocabulary were late age of onset to Swedish and/or a low daily language exposure to Swedish. The Swedish vocabulary scores and the effect of background factors are analysed statistically in Section 6.3.3.6.

\textsuperscript{153} No information about language exposure and language development patterns was available in the questionnaire for BiAra7-05. However, during the Swedish testing session, BiAra7-05 mentioned to the experimenter that he had recently arrived to Sweden. As for BiAra7-16, the parents reported that the child had only been exposed to Swedish for about a year.

\textsuperscript{154} It should be noted, however, that BiAra6-26 was quite restless and not fully cooperating with the experimenters during both Arabic and Swedish testing sessions.

\textsuperscript{155} Despite having been born in Sweden, BiAra6-15 was reported to have an onset of Swedish at the age of 2 but was only frequently exposed to Swedish from the age of 6 (i.e. when the child was enrolled in preschool).
6.3.2.3 Comparison of Arabic and Swedish vocabulary scores and age development

In this section, Arabic and Swedish scores are compared to each other for both comprehension and production in an aim to observe whether the children’s performance differed in the two languages and to test whether there was a correlation between the scores of the two languages.

Figure 6.7 shows the Arabic and Swedish comprehension scores plotted against each other for all age groups.

![Figure 6.7. Scatterplot of Arabic (CLT) comprehension scores (y-axis) and Swedish (CLT) comprehension scores (x-axis). The grey lines indicate a 50% mark (30 points) of the max score.](image)

As seen in Figure 6.7, the majority of the children are found in the top right quadrant, meaning that they scored more than 50% (30 points) on both Arabic and Swedish (CLT) comprehension. In other words, children who had high comprehension scores in Arabic also had high comprehension scores in Swedish, and vice versa.

A Pearson correlation (two-tailed) revealed that the Arabic and Swedish vocabulary (CLT) comprehension scores indeed correlated with each other but that the correlation was weak to medium \( r = .293, N = 100, p = .003 \).

Only one 4-year-old child (BiAra4-16, who has already been discussed in Section 6.3.2.1) is found in the bottom left quadrant, meaning that this child scored below 50% in both Arabic and Swedish comprehension. Some children have a higher comprehension score in one language but below average (above 50% score) in the other language. Children found in the bottom right quadrant (BiAra4-02, BiAra4-15, and BiAra5-06) scored better in Swedish comprehension than Arabic comprehension, whereas ten children, from all age groups,
found in the top left quadrant, scored better in Arabic comprehension than Swedish comprehension. There are more children in the top left quadrant than in the bottom right one, indicating that more children seem to have stronger receptive vocabulary skills in Arabic than in Swedish (but the majority of the children seem to have a balance between the receptive vocabularies in their two languages).

Let us now move on to vocabulary production. Figure 6.8 shows the Arabic and Swedish production scores plotted against each other for all age groups.

![Figure 6.8. Scatterplot of Arabic (CLT) production scores (y-axis) and Swedish (CLT) production scores (x-axis). The grey lines indicate a 50% mark (30 points) of the max score.](image)

Unlike the comprehension scores, Figure 6.8 shows how much more scattered the production scores are. Indeed, a Pearson correlation (two-tailed) showed no significant correlation between the Arabic and Swedish production scores ($r = -0.113, N = 100, p = 0.264$). Children who are found in the upper right quadrant performed above 50% in both of their languages in production. The majority of these children belong to the oldest age groups. Children in the bottom left quadrant performed low in both Arabic and Swedish production; the majority of these children are 4-year-olds. Children who are found in the top-left or bottom-right quadrants are considered ‘Arabic dominant’ or ‘Swedish dominant’, respectively, since they have an uneven distribution of their expressive vocabulary skills in both languages. This pattern is very different than that in Figure 6.7 (comprehension scores) where the majority of the children gathered in the top right quadrant, indicating that they scored above the average (above 50% score) in both Arabic comprehension and Swedish comprehension.
What follows is a comparison of the means of Arabic and Swedish comprehension and production scores over age. Figure 6.9 visualizes the mean scores of Arabic and Swedish comprehension and production for each age group.

As shown in Figure 6.9, comprehension scores are higher than production scores for both languages (also recall Section 6.3.1). The distance between comprehension and production means remains nearly constant for both languages across all age groups. The clearest difference in mean scores is between the 4-year-olds and the 7-year-olds for both comprehension and production in both languages.

The mean scores for Arabic comprehension and production remain higher than the Swedish ones throughout the 4-year-old to the 6-year-old group. However, both Swedish comprehension and production scores for the 7-year-old group are slightly higher than the Arabic scores. In general, the development for comprehension and production is steeper in Swedish than in Arabic. The error bars are larger for the production scores in both languages, thus visualizing the great individual variation in the production scores across ages.

As for the difference between the two languages within each age group, paired samples t-tests showed that there was no significant difference between Arabic and Swedish comprehension and production scores, except at age 4 for the comprehension scores. The 4-year-olds had significantly higher Arabic comprehension scores ($t(21) = 2.332, p = .030, d = .497, \text{small to moderate effect size}$) than Swedish comprehension scores. This difference between the two languages was not significant for any other age group or for any other language measure, neither comprehension nor production.
6.3.2.4 Vocabulary scores and age: summary
For both Arabic and Swedish, comprehension scores were higher than the production scores. Also, both comprehension and production scores in both languages correlated with each other as well as increased significantly with age (in months). The two languages’ comprehension scores correlated with each other while the production scores did not. Individual children’s background factors were discussed in relation to their low scores where the children’s daily language exposure pattern seems to stand out as a contributing factor for the low scores in the respective language. The following section investigates the relationship between background factors and vocabulary scores using statistical methods.

6.3.3 Vocabulary scores and background factors
In this section, Arabic and Swedish (CLT) comprehension and production scores are explored in relation to various background factors. The section starts with the exploration of Arabic and Swedish scores in relation to each of the background factors: age of onset of Swedish (6.3.3.1), socio-economic status (6.3.3.2), language use in the home (6.3.3.3), language use outside the home (6.3.3.4), daily language exposure (6.3.3.5). Then, four regression analyses (one for each vocabulary and language measure) are presented (6.3.3.6). The section ends with a summary of the results (6.3.3.7).

6.3.3.1 Age of onset (Swedish)
Since age of onset (AoO) of Arabic was almost identical for all children ($N = 98$), i.e. from birth, no statistical analyses were made on this variable. The children’s AoO to Swedish varied considerably; hence, the children were divided into two groups: children who had started hearing Swedish before the age of 3 ($N = 48$) and those who had started hearing Swedish at the age of 3 or after ($N = 51$). There was a relatively even distribution of AoO of Swedish among the age groups.\footnote{The distribution of AoO of Swedish in the age groups was as follows: 4-year-olds: before age 3 $N = 15$, after age 3 $N = 7$; 5-year-olds: before age 3 $N = 12$, after age 3 $N = 13$; 6-year-olds: before age 3 $N = 13$, after age 3 $N = 16$; 7-year-olds: AoO before age 3 $N = 8$, after age 3 $N = 15$.}

Table 6.4 shows the mean, standard deviation (SD) and range of Arabic and Swedish vocabulary scores according to the children’s AoO of Swedish. A noticeable difference of approximately 10 points can be observed in the Arabic production scores between the children whose AoO of Swedish was before age 3 and after age 3. The range of scores differed between these groups as well (1–48, 12–53).
Table 6.4. Age of onset (AoO) of Swedish and Arabic and Swedish vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th></th>
<th>AoO of Swedish before age 3 (N = 48)</th>
<th>AoO of Swedish after age 3 (N = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>45.0</td>
<td>49.8</td>
</tr>
<tr>
<td>SD</td>
<td>7.7</td>
<td>6.6</td>
</tr>
<tr>
<td>Range</td>
<td>25–57</td>
<td>26–59</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>27.0</td>
<td>37.9</td>
</tr>
<tr>
<td>SD</td>
<td>13.5</td>
<td>8.2</td>
</tr>
<tr>
<td>Range</td>
<td>1–48</td>
<td>12–53</td>
</tr>
<tr>
<td>Swedish comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>48.7</td>
<td>43.0</td>
</tr>
<tr>
<td>SD</td>
<td>9.9</td>
<td>10.7</td>
</tr>
<tr>
<td>Range</td>
<td>18–60</td>
<td>23–60</td>
</tr>
<tr>
<td>Swedish production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>34.6</td>
<td>27.6</td>
</tr>
<tr>
<td>SD</td>
<td>11.3</td>
<td>11.0</td>
</tr>
<tr>
<td>Range</td>
<td>16–53</td>
<td>10–52</td>
</tr>
</tbody>
</table>

Note. For each language, max comprehension score = 60 points, max production score = 60 points.

Independent samples t-tests showed that children who had regular exposure to Swedish (in addition to Arabic) before the age of 3 had significantly higher Swedish comprehension ($t(96.98) = 2.735, p = .007, d = .549$, moderate effect size) and production scores ($t(96.23) = 3.215, p = .002, d = .629$, moderate effect size) but also significantly lower Arabic comprehension ($t(92.81) = -3.283, p = .001, d = -.663$, moderate effect size) and production scores ($t(76.91) = -4.801, p < .001, d = -.979$, large effect size).

6.3.3.2 SES (parental education)

Parental education was used as a proxy for socio-economic status (SES). In order to investigate whether SES had an impact on vocabulary, the children were divided into two groups: children whose parents had ‘low-SES’ (no tertiary education, ISCED-levels 0–3, $N = 30$) and children whose parents had ‘high-SES’ (at least tertiary education achieved, ISCED-levels 4–8, $N = 64$).\(^{157}\)

Independent samples t-tests showed that for Arabic comprehension there was no significant difference between the groups ($t(47.61) = .232, p = .818$). However, for Arabic production, a difference emerged between the SES groups ($t(62.19) = 2.214, p = .031$) where children of high-SES parents scored significantly lower in Arabic production than children of low-SES parents.

\(^{157}\) There was an even distribution of SES across all age groups.
Effect size was small though \((d = .473)\). As for the Swedish vocabulary scores, no significant difference between the SES groups was found (Swedish: comprehension \((t(59.16) = -1.833, p = .072)\); production \((t(60.87) = -1.465, p = .148)\). \[158\]

| Table 6.5. SES (two-way split) in relation to Arabic and Swedish vocabulary (CLT) scores. |
|-----------------|-----------------|
| Arabic comprehension | Arabic production |
| **Mean** | **Mean** |
| 48.1 | 36.7 |
| **SD** | **SD** |
| 8.2 | 10.8 |
| **Range** | **Range** |
| 25–59 | 12–53 |

**Swedish comprehension**

| **Mean** | **Mean** |
| 43.8 | 47.9 |
| **SD** | **SD** |
| 9.8 | 10.2 |
| **Range** | **Range** |
| 26–60 | 18–60 |

**Swedish production**

| **Mean** | **Mean** |
| 29.2 | 32.8 |
| **SD** | **SD** |
| 10.8 | 11.7 |
| **Range** | **Range** |
| 15–52 | 10–53 |

*Note. For each language, max comprehension score = 60 points, max production score = 60 points.*

6.3.3.3 Language use in the home

Concerning the parents’ language use with the child, the children were classified into two categories: ‘only/mostly Arabic’, if *both* parents spoke ‘(almost) only Arabic’ or ‘mostly Arabic, sometimes Swedish’ to the child \((N = 79)\), and ‘other’ if parents used any other language combination at home \((N = 20)\). \[160\]

---

[158] Note that some SES values in the present analysis differ slightly from those reported in Bohnacker et al. (2021) due to an updated analyses of the parents’ education levels.

[159] For further exploration, SES was also cut into three groups (low-SES for ISCED 0–3, \(N = 30\); mid-SES for ISCED 4–5, \(N = 33\), high-SES for ISCED 6–8, \(N = 33\)) where one-way ANOVAs did not show any significant difference between the groups on any vocabulary measure (Arabic comprehension: \(F(2,97) = .1061, p = .896\); Arabic production: \(F(2,97) = 2.925, p = .058\); Swedish comprehension: \(F(2,97) = .866, p = .424\); Swedish production: \(F(2,97) = .487, p = .616\)). When SES was treated as a continuous variable ranging from 0–8 on the ISCED scale, Pearson correlations (two-tailed) showed no significant correlation between SES and any vocabulary measure either (Arabic comprehension \(p = .896\); Arabic production \(p = .135\); Swedish comprehension \(p = .083\), Swedish production \(p = .149\)).

[160] The two categories ‘Only/mostly Arabic’ and ‘Other’ were evenly distributed across the age groups.
Table 6.6. Parents’ language use with child in relation to Arabic and Swedish vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th>Language Use</th>
<th>Only/mostly Arabic (N = 79)</th>
<th>Other (N = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>49.0</td>
<td>41.6</td>
</tr>
<tr>
<td>SD</td>
<td>6.3</td>
<td>9.2</td>
</tr>
<tr>
<td>Range</td>
<td>29–59</td>
<td>25–54</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>34.9</td>
<td>23.4</td>
</tr>
<tr>
<td>SD</td>
<td>11.5</td>
<td>11.2</td>
</tr>
<tr>
<td>Range</td>
<td>1–53</td>
<td>10–42</td>
</tr>
<tr>
<td>Swedish comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>45.1</td>
<td>48.2</td>
</tr>
<tr>
<td>SD</td>
<td>10.7</td>
<td>10.6</td>
</tr>
<tr>
<td>Range</td>
<td>18–60</td>
<td>29–60</td>
</tr>
<tr>
<td>Swedish production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>29.8</td>
<td>35.9</td>
</tr>
<tr>
<td>SD</td>
<td>11.3</td>
<td>11.9</td>
</tr>
<tr>
<td>Range</td>
<td>10–53</td>
<td>17–53</td>
</tr>
</tbody>
</table>

Note. For each language, max comprehension score = 60 points, max production score = 60 points.

Independent samples t-tests revealed that children whose parents spoke to them ‘only/mostly in Arabic’ scored significantly higher in both Arabic comprehension ($t(23.65) = 3.420, p = .002, d = 1.071$) and Arabic production ($t(30.05) = 4.115, p < .001, d = 1.012$) than children whose parents did not. As for the Swedish scores, children whose parents spoke to them ‘only/mostly Arabic’ did not score lower in Swedish comprehension than other children ($t(29.514) = -1.158, p = .256$) but they scored significantly lower in Swedish production ($t(28.34) = -2.078, p = .047, d = -.536, moderate effect size$). Note though that the $p$-value only just reached significance ($p = .047$).

In order to investigate whether the child’s language use with the parents was significantly related to vocabulary, children were divided into two groups; whether they spoke ‘only/mostly Arabic’ ($N = 68$) to both of their parents, or whether they spoke any ‘other’ language combination ($N = 31$).161

Independent samples t-tests revealed that the languages used by the child with the parents have a clear significant effect on all four vocabulary measures (Arabic comprehension: $t(46.10) = 3.108, p = .003, d = .748$, moderate effect size; Arabic production: $t(45.80) = 6.520, p < .001, d = 1.413$, large effect size; Swedish comprehension: $t(75.77) = -4.273, p < .001, d = -.926$, large effect size; Swedish production: $t(61.60) = -5.621, p < .001, d = -1.190$, large effect size).

161 For the independent variable ‘child’s language use with the parents’, the two categories ‘Only/mostly Arabic’ and ‘Other’ were evenly distributed across the age groups.
Table 6.7. Child’s language use with parents in relation to Arabic and Swedish vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th></th>
<th>Only/mostly Arabic (N = 68)</th>
<th>Other (N = 31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>49.1</td>
<td>43.8</td>
</tr>
<tr>
<td>SD</td>
<td>6.4</td>
<td>8.5</td>
</tr>
<tr>
<td>Range</td>
<td>26–59</td>
<td>25–56</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>37.2</td>
<td>22.6</td>
</tr>
<tr>
<td>SD</td>
<td>9.24</td>
<td>12.4</td>
</tr>
<tr>
<td>Range</td>
<td>12–53</td>
<td>1–45</td>
</tr>
<tr>
<td>Swedish comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>42.9</td>
<td>52.0</td>
</tr>
<tr>
<td>SD</td>
<td>10.6</td>
<td>8.0</td>
</tr>
<tr>
<td>Range</td>
<td>18–60</td>
<td>33–60</td>
</tr>
<tr>
<td>Swedish production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>27.2</td>
<td>39.4</td>
</tr>
<tr>
<td>SD</td>
<td>10.4</td>
<td>9.8</td>
</tr>
<tr>
<td>Range</td>
<td>10–52</td>
<td>19–53</td>
</tr>
</tbody>
</table>

Note. For each language, max comprehension score = 60 points, max production score = 60 points.

As for the child’s language use with sibling(s), two different split comparisons were made. In the first comparison, children who spoke ‘mostly Arabic’ (N = 39) were compared with all other children who spoke other language combinations with their sibling(s) (N = 54). Independent samples t-tests revealed that children who spoke ‘mostly Arabic’ with their siblings had both significantly higher Arabic scores (Arabic comprehension: t(90.09) = 2.273, p = .025, d = .460, small effect size; Arabic production: t(89.44) = 4.410, p < .001, d = .927, large effect size), and significantly lower Swedish scores (Swedish comprehension: t(77.87) = -3.308, p = .001, d = -.705, moderate effect size; Swedish production: t(88.41) = -3.824, p < .001, d = -.783, moderate effect size) than children who spoke ‘mostly Swedish’, ‘both Arabic and Swedish’ or an ‘other’ language combination (for example, with an addition of English).

When the second comparison was made regarding whether the child spoke ‘mostly Swedish’ with their sibling(s) (N = 18) or ‘other’ language combination(s) (N = 75), independent samples t-tests revealed that children who spoke ‘only/mostly Swedish’ with their siblings scored significantly lower in Arabic production (t(25.31) = -4.226, p < .001, d = -1.128, large effect size) and significantly higher in Swedish scores (comprehension: t(36.19) = 4.492, p < .001, d = .945, large effect size; production: t(28.11) = 5.010, p < .001, d = .945).

162 Both comparison splits regarding children’s language use with their sibling(s) (‘mostly Arabic’ vs. ‘other’ and ‘mostly Swedish’ vs. ‘other’) had an even distribution across the age groups.
1.229, large effect size) than children who spoke ‘other’ language combinations with their sibling(s). As for Arabic comprehension scores, there was no significant difference between the two groups \( (t(22.24) = -1.866, p = .075) \).

In order to investigate whether children who were the oldest sibling (first born, \( N = 33 \)) scored significantly higher on vocabulary measures than children who had older siblings (\( N = 61 \)) did, independent samples t-tests were run.

**Table 6.8.** Children’s Arabic and Swedish vocabulary (CLT) scores in relation to birth order.

<table>
<thead>
<tr>
<th></th>
<th>First born (( N = 33 ))</th>
<th>Has older sibling(s) (( N = 61 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>47.6</td>
<td>47.4</td>
</tr>
<tr>
<td>SD</td>
<td>8.0</td>
<td>7.4</td>
</tr>
<tr>
<td>Range</td>
<td>27–57</td>
<td>25–59</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>33.7</td>
<td>32.5</td>
</tr>
<tr>
<td>SD</td>
<td>12.1</td>
<td>12.2</td>
</tr>
<tr>
<td>Range</td>
<td>10–51</td>
<td>1–53</td>
</tr>
<tr>
<td>Swedish comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>45.2</td>
<td>46.5</td>
</tr>
<tr>
<td>SD</td>
<td>19.9</td>
<td>10.3</td>
</tr>
<tr>
<td>Range</td>
<td>27–60</td>
<td>26–60</td>
</tr>
<tr>
<td>Swedish production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>29.5</td>
<td>32.0</td>
</tr>
<tr>
<td>SD</td>
<td>11.1</td>
<td>12.2</td>
</tr>
<tr>
<td>Range</td>
<td>10–52</td>
<td>11–53</td>
</tr>
</tbody>
</table>

*Note.* For each language, max comprehension score = 60 points, max production score = 60 points.

The results showed no significant difference between the two groups for any vocabulary measure in any language (Arabic: comprehension \( t(61.10) = .117, p = .908 \); production \( t(66.02) = .460, p = .647 \); Swedish: comprehension \( t(62.67) = -557, p = .579 \); production \( t(71.55) = -1.000, p = .321 \)). Thus, being the older sibling was no advantage concerning vocabulary results.

Another way to understand language use in the home is to observe the language-stimulating activities that parents are performing with their children. Joint book reading is believed to particularly boost children’s vocabulary knowledge (and narrative competence). Parents’ responses in the questionnaire were divided into two categories for each language. For Arabic vocabulary scores: ‘Reading in Arabic often’ (\( N = 53 \)), ‘Reading in Arabic rarely’ (\( N = 25 \)); for Swedish vocabulary scores: ‘Reading is Swedish often’ (\( N = 49 \)) and ‘Reading in Swedish rarely’ (\( N = 34 \)).
Table 6.9. Children’s Arabic and Swedish vocabulary (CLT) scores in relation to joint book reading.

<table>
<thead>
<tr>
<th></th>
<th>Joint reading in Arabic often (N = 53)</th>
<th>Joint reading in Arabic rarely (N = 25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>49.6</td>
<td>44.2</td>
</tr>
<tr>
<td>SD</td>
<td>5.9</td>
<td>8.1</td>
</tr>
<tr>
<td>Range</td>
<td>25–58</td>
<td>26–56</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>37.0</td>
<td>27.5</td>
</tr>
<tr>
<td>SD</td>
<td>10.1</td>
<td>12.4</td>
</tr>
<tr>
<td>Range</td>
<td>11–53</td>
<td>9–47</td>
</tr>
<tr>
<td>Swedish comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>49.0</td>
<td>41.8</td>
</tr>
<tr>
<td>SD</td>
<td>10.0</td>
<td>9.9</td>
</tr>
<tr>
<td>Range</td>
<td>26–60</td>
<td>18–58</td>
</tr>
<tr>
<td>Swedish production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>34.1</td>
<td>26.6</td>
</tr>
<tr>
<td>SD</td>
<td>11.8</td>
<td>9.6</td>
</tr>
<tr>
<td>Range</td>
<td>11–53</td>
<td>10–47</td>
</tr>
</tbody>
</table>

Note. For each language, max comprehension score = 60 points, max production score = 60 points.

Independent samples t-tests revealed that children whose parents read with them in Arabic had significantly higher Arabic comprehension ($t(36.18) = -2.992, p = .005, d = -.815$, large effect size) and Arabic production scores ($t(39.50) = -3.349, p = .002, d = -.876$, large effect size) than children who did not perform such activities.

Similarly, children who performed joint reading activities in Swedish with their parents at home had significantly higher Swedish comprehension ($t(71.52) = -3.241, p = .002, d = -.772$, moderate effect size) and Swedish production scores ($t(78.72) = -3.153, p = .002, d = -.679$, moderate effect size) than children whose parents did not read with them.

However, note that the age distribution for both reading activities (whether in Arabic or Swedish) was not even, as the vast majority of the 6-year-olds and almost all of the 7-year-olds performed joint reading activities often (with their parents) at home, whilst this was not the case for the younger two age groups.163

---

163 The distribution of the children across the age groups was not even for ‘joint reading in Arabic often/rarely’: 4-year-olds: often N = 10, rarely N = 7; 5-year-olds: often N = 13, rarely N = 9; 6-year-olds: often N = 15, rarely N = 7; 7-year-olds: often N = 15, rarely N = 2. The distribution of the children across the age groups was not even for ‘joint reading in Swedish...
6.3.3.4 Home language use outside the home

In order to investigate the child’s home language use outside the home, parents were asked to answer ‘yes-or-no’ questions on whether their child heard the home language (Arabic) from extended family members, friends/playmates and staff members at school.

Regarding home language use with extended family members, independent samples t-tests showed that children who heard Arabic from extended family members ($N = 78$) did not score significantly differently in their Arabic comprehension ($t(31.30) = -0.283, p = .779$) or production ($t(33.36) = 1.027, p = .312$) than other children ($N = 21$).

The children who heard Arabic from friends or playmates ($N = 29$) were compared to those who did not receive such input ($N = 70$). Independent samples t-tests revealed no significant difference between the groups regarding their scores in Arabic comprehension ($t(75.20) = -1.811, p = .074$) but a significant effect on Arabic production ($t(69.94) = -2.011, p = .048, d = -0.395$, small effect size).

Some children were reported to hear the home language from Arabic-speaking staff members ($N = 25$) whilst the majority of the children did not ($N = 72$). There was no significant difference between the groups for Arabic comprehension ($t(43.72) = -0.253, p = .801$) and production ($t(47.60) = -1.023, p = .311$). Note, however, that there was not an even distribution across the age groups since the majority of the children who had Arabic-speaking staff members belonged to the younger age groups.

The effect of attending mother tongue instruction (MTI) was also measured via independent samples t-tests: children who attended MTI classes ($N = 65$) were compared to children who did not ($N = 31$). Children who attended MTI classes scored significantly higher than non-attendees in Arabic comprehension ($t(45.43) = -3.738, p = .001, d = -.914$, large effect size) and production ($t(48.09) = -3.563, p = .001, d = -.848$, large effect size). Note that there was no even age distribution for this variable since the majority of the children who attended MTI belonged to the two older age groups.

often/rarely’ either: 4-year-olds: often $N = 10$, rarely $N = 10$; 5-year-olds: often $N = 9$, rarely $N = 14$; 6-year-olds: often $N = 13$, rarely $N = 8$; 7-year-olds: often $N = 17$, rarely $N = 2$.

The distribution of the children for ‘home language from extended family’ (yes/no) was not even across the age groups: 4-year-olds: yes $N = 14$, no $N = 8$; 5-year-olds: yes $N = 20$, no $N = 5$; 6-year-olds: yes $N = 23$, no $N = 6$; 7-year-olds: yes $N = 21$, no $N = 2$, i.e. the vast majority of the older children heard Arabic from extended family. What is not known, however, is the quantity and quality of this Arabic input.

The distribution of the children for 'home language from friends or playmates' (yes/no) was relatively even across age groups.

The distribution of the children across the age groups was not even for ‘hearing home language from school staff’ (yes/no): 4-year-olds: yes $N = 7$, no $N = 15$; 5-year-olds: yes $N = 9$, no $N = 15$; 6-year-olds: yes $N = 5$, no $N = 23$; 7-year-olds: yes $N = 4$, no $N = 19$.

The distribution of the children across the age groups was not even for ‘attending MTI’ (yes/no): 4-year-olds: yes $N = 10$, no $N = 10$; 5-year-olds: yes $N = 10$, no $N = 14$; 6-year-olds: yes $N = 23$, no $N = 6$; 7-year-olds: yes $N = 22$, no $N = 1$.
6.3.3.5 Daily language exposure

Children’s current daily language exposure was estimated by the parents: about one third of the children (N = 30) were receiving the majority of their daily input in Arabic (60%–90%), a little over one third (N = 37) had an even exposure to Arabic and Swedish (50%–50%), and around a third (N = 30) had the majority of their daily language input in Swedish (60%–90%). The distribution of the children across the age groups was not even: many of the younger children had more exposure to Arabic and many of the older children were evenly exposed to both languages or had more exposure to Swedish.168

One-way ANOVAs showed a significant effect of the daily language exposure pattern. Children who received the majority of their input in Arabic (‘mostly Arabic’) had higher Arabic comprehension \( (F(2,94) = 3.108, p = .049, \eta^2 = .062, \text{moderate effect size}) \) and production scores \( (F(2,94) = 13.562, p < .001, \eta^2 = .224, \text{large effect size}) \). Comparing the significance (p-values) and the effect sizes (\( \eta^2 \) values), having a daily language exposure of ‘mostly Arabic’ seems to have a marginal effect on Arabic comprehension but a clear effect on Arabic production.

Children who received ‘mostly Swedish’ input had higher Swedish comprehension \( (F(2,94) = 10.972, p < .001, \eta^2 = .189, \text{large effect size}) \) and production scores \( (F(2,94) = 15.317, p < .001, \eta^2 = .246, \text{large effect size}) \). Having a daily language exposure to ‘mostly Swedish’ had a clear effect on both Swedish comprehension and Swedish production. The eta-squared (\( \eta^2 \)) values suggest that the effects of exposure in both languages are stronger for vocabulary production than for comprehension.

Post hoc analyses (with Bonferroni correction) revealed that for Arabic comprehension, there was no significant difference between any of the three exposure groups (‘Mostly Arabic’ vs. ‘Even’ \( p = 1.000 \), ‘Mostly Arabic’ vs. ‘Mostly Swedish’ \( p = .078 \), ‘Even’ vs. ‘Mostly Swedish’ \( p = .121 \)). For Arabic production, a significant difference emerged between the ‘Mostly Swedish’ and the remaining two groups (‘Mostly Swedish’ vs. ‘Mostly Arabic’ \( p < .001 \), ‘Mostly Swedish’ vs. ‘Even’ \( p < .001 \)) but not between ‘Even’ vs. ‘Mostly Arabic’ \( (p = .927) \). This means that children who had more Swedish exposure throughout the day had significantly lower Arabic production scores than children who had an even exposure to both languages or had more daily exposure to Arabic. Figure 6.10 visualizes the distribution of the Arabic comprehension scores with respect to estimated daily language exposure, while Figure 6.11 visualizes the distribution of the Arabic production scores. The outliers in both figures represent two 4-year-olds, namely BiAra4-16 in the ‘Mostly Arabic’ and BiAra4-02 in the ‘Even’ boxplot.

---

168 For ‘estimated daily language exposure’, the distribution across the age groups was not even:

- 4-year-olds: Mostly Arabic N = 9, Even N = 7, Mostly Swedish N = 6;
- 5-year-olds: Mostly Arabic N = 11, Even N = 7, Mostly Swedish N = 7;
- 6-year-olds: Mostly Arabic N = 5, Even N = 13, Mostly Swedish N = 11;
- 7-year-olds: Mostly Arabic N = 5, Even N = 10, Mostly Swedish N = 6, Other N = 2.
Table 6.10. Current daily language exposure in relation to Arabic and Swedish vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th></th>
<th>Mostly Arabic (N = 30)</th>
<th>Even exposure (N = 37)</th>
<th>Mostly Swedish (N = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>49.0</td>
<td>48.5</td>
<td>44.7</td>
</tr>
<tr>
<td>SD</td>
<td>6.8</td>
<td>6.1</td>
<td>9.3</td>
</tr>
<tr>
<td>Range</td>
<td>26–59</td>
<td>29–58</td>
<td>25–58</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>38.0</td>
<td>35.2</td>
<td>24.0</td>
</tr>
<tr>
<td>SD</td>
<td>8.4</td>
<td>11.0</td>
<td>13.2</td>
</tr>
<tr>
<td>Range</td>
<td>12–52</td>
<td>1–53</td>
<td>8–52</td>
</tr>
<tr>
<td>Swedish comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>39.1</td>
<td>46.8</td>
<td>50.6</td>
</tr>
<tr>
<td>SD</td>
<td>9.8</td>
<td>9.5</td>
<td>9.9</td>
</tr>
<tr>
<td>Range</td>
<td>18–59</td>
<td>28–60</td>
<td>23–60</td>
</tr>
<tr>
<td>Swedish production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>23.5</td>
<td>31.4</td>
<td>37.9</td>
</tr>
<tr>
<td>SD</td>
<td>8.5</td>
<td>10.8</td>
<td>10.7</td>
</tr>
<tr>
<td>Range</td>
<td>10–42</td>
<td>12–52</td>
<td>14–53</td>
</tr>
</tbody>
</table>

Note. For each language, max comprehension score = 60 points, max production score = 60 points.

Figure 6.10. Box plots of estimated daily language exposure and Arabic comprehension scores. The dots indicate outliers. Error bars indicate a Confidence Interval (CI) of 95%.
Let us now move on to daily language exposure and Swedish vocabulary. As for the Swedish *comprehension* scores, no significant difference was found between ‘Even’ vs. ‘Mostly Swedish’ ($p = .337$) but rather between ‘Mostly Arabic’ vs. ‘Mostly Swedish’ ($p < .001$), and between ‘Even’ vs. ‘Mostly Arabic’ ($p = .005$). Finally, for the Swedish *production* scores, the post hoc analyses (with Bonferroni correction) revealed a significant difference between all groups (‘Mostly Arabic’ vs. Even $p = .006$, ‘Mostly Arabic’ vs. ‘Mostly Swedish’ $p < .001$, ‘Mostly Swedish’ vs. ‘Even’ $p = .030$). Figure 6.12 visualizes the distribution of the Swedish comprehension scores with respect to current daily language exposure. The outliers in the ‘Mostly Swedish’ box plot represent one 4-year-old (BiAra4-03, lowest outlier) and one 6-year-old (BiAra6-26, the higher outlier). Figure 6.13 visualizes the distribution of the Swedish production scores with respect to current daily language exposure.
Another way to visualize the effect of the daily language exposure on the Arabic vocabulary scores is to plot the children’s individual scores by age while specifying each child’s estimated daily exposure as well as the variety of Arabic. The specification of each child’s variety in the figures gives an idea of whether groups of children who performed similarly spoke the same Arabic variety. Figure 6.14 shows the children’s Arabic comprehension scores and Figure 6.15 shows the children’s Arabic production scores plotted against age (in months) where the shape of the dot indicates the Arabic variety of the child (Egyptian, Iraqi, Lebanese, Palestinian, Syrian), and the colour of the dot indicates daily language exposure.
Figure 6.14. Scatterplot of Arabic comprehension scores and age (in months) in relation to estimated daily language exposure and Arabic variety (Egy = Egyptian, Irq = Iraqi, Leb = Lebanese, Pal = Palestinian, Syr = Syrian, Maj Ara = Majority in Arabic, 50/50 = Even exposure, Maj Swe = Majority in Swedish). The figure was created by Linnéa Öberg.

Figure 6.15. Scatterplot of Arabic production scores and age (in months) in relation to estimated daily language exposure and Arabic variety (Egy = Egyptian, Irq = Iraqi, Leb = Lebanese, Pal = Palestinian, Syr = Syrian, Maj Ara = Majority in Arabic, 50/50 = Even exposure, Maj Swe = Majority in Swedish). This figure was previously published in Bohnacker, Haddad & Öberg (2021) and created by Linnéa Öberg.

The fact that the Arabic varieties are distributed evenly across the scatterplot suggests that the Arabic vocabulary (CLT) comprehension task does not disadvantage any particular Arabic variety group. Furthermore, the majority of the low scoring children were reported to have the majority of their daily language exposure in Swedish, as is visualized by the yellow colour. The low
scoring child with the blue circle (i.e. a child who speaks a Syrian Arabic variety and has the majority of his daily language exposure in Arabic) is the 4-year-old child (BiAra4-16) who was reported to have had additional language exposure to English at home. The low scoring child who is visualized with a green triangle is a speaker of Lebanese Arabic and was reported to be exposed evenly to both languages. This child, BiAra4-02, has been discussed earlier and did not cooperate well during the Arabic testing session. Furthermore, the child’s parent mentions in the parental interview (in the longitudinal study) that the child has always favoured speaking Swedish over Arabic. The remaining four low-scoring (yellow diamonds) were Iraqi Arabic-speaking children who were reported to have the majority of their daily language exposure in Swedish.

Similarly to the Arabic comprehension scores, the Arabic production scores are distributed relatively evenly with regard to Arabic variety across the scatterplot, suggesting that the Arabic vocabulary (CLT) production task does not disadvantage any particular Arabic variety group. In other words, when the children’s production answers were scored, no variety appears to have been favoured over the other. The estimated daily exposure colour code points to the strong effect of language exposure on Arabic production scores. The majority of the low scoring children were reported to have the majority of their daily language exposure in Swedish (yellow coloured shapes). The lowest scoring green coloured triangle (i.e. the 4-year-old Lebanese Arabic-speaking child is BiAra4-02) was mentioned above.

6.3.3.6 Multivariate analyses

In this section, the combined effect of various background factors on the four vocabulary measures (Arabic comprehension, Arabic production, Swedish comprehension and Swedish production) is investigated separately. Since there was a significant increase with age for all vocabulary measures, chronological age (in months) was also included in the regression analyses as a control variable. For each of the four linear regression models, only those background factors that had individually proven to be significant for the vocabulary measure in question were included in the model. Hence, for Arabic comprehension, age (in months), AoO of Swedish, parents’ language to the child, joint reading in Arabic, MTI attendance and daily language exposure in Swedish were added to the regression model. For Arabic production, the same variables as for Arabic comprehension were included in the regression model with the addition of SES (binary split). For Swedish comprehension, age (in months), AoO of Swedish, joint reading in Swedish and daily language exposure in Arabic were included in the regression model. For Swedish production, the same variables as Swedish comprehension were included, with the addition of parents’ language use with the child. All independent variables in the four regression analyses were binary except for age, which was a continuous variable. The variable AoO of Swedish was included in both the Swedish and
Arabic regression analyses to investigate whether AoO of Swedish had an effect on the vocabulary measures in combination with the remaining independent variables.

The regression models included the variable ‘daily language exposure’ to test whether overall daily language exposure in a certain language affected the vocabulary scores in the other language. Note that if we were to include daily language exposure in Arabic together with other Arabic input variables, this would not show much in the regression analyses for Arabic due to high collinearity between these variables. Therefore, ‘mostly Swedish daily language exposure’ was added as a variable to the regression models for Arabic comprehension and Arabic production to see whether Swedish exposure had a negative effect on Arabic vocabulary. Conversely, ‘mostly Arabic daily language exposure’ was added to the analyses for Swedish comprehension and Swedish production. In order to do this, the variable ‘daily language exposure’, which was divided into 3 categories when analysed individually (one way ANOVA, in Section 6.3.3.5), was recoded into three binary variables (‘mostly Arabic’ vs. ‘other’; ‘even’ vs. ‘other’; ‘mostly Swedish’ vs. ‘other’).

Background variables that were related to the child’s language output were not included in the regression analyses. These variables are: child language use with parents, extended family members, friends/playmates and staff members. In what follows, four multiple linear regression analyses are presented, one for each language and vocabulary measure.

6.3.3.6.1 Multiple regression models for Arabic vocabulary scores

A multiple regression was calculated to predict Arabic comprehension scores based on the background factors that proved significant individually. Arabic comprehension was the dependent variable and age (months), AoO of Swedish, parents’ L1 (Arabic) language with child, joint reading in Arabic, attending MTI and daily language exposure in Swedish were the independent variables. Table 6.11 presents the regression model for Arabic comprehension.

Table 6.11. Linear regression model of the combined effect on Arabic (CLT) comprehension scores.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>29.23</td>
<td>3.15</td>
<td>.42</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>Age</td>
<td>.22</td>
<td>.05</td>
<td>.42</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>AoO of Swedish</td>
<td>1.01</td>
<td>1.21</td>
<td>.07</td>
<td>.408</td>
</tr>
<tr>
<td>Parents’ L1 language with child</td>
<td>7.14</td>
<td>1.58</td>
<td>.39</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>Joint reading in Arabic</td>
<td>2.26</td>
<td>1.34</td>
<td>.15</td>
<td>.096</td>
</tr>
<tr>
<td>MTI attendance</td>
<td>2.55</td>
<td>1.40</td>
<td>.16</td>
<td>.073</td>
</tr>
<tr>
<td>Daily exposure to Swedish</td>
<td>-.77</td>
<td>1.41</td>
<td>-.05</td>
<td>.587</td>
</tr>
</tbody>
</table>

R² (adjusted) = .549

Note. * = p < .05, ** = p < .01, *** = p < .001. F(6,69) = 16.23, p < .001.
The results of the regression showed that the model explained 54.9% of the variance in the Arabic comprehension scores, with only age and parents’ L1 language to the child being significant predictors. There was no significant effect of AoO of Swedish, attending MTI, joint reading in Arabic or daily exposure in Swedish. This means that having early exposure to Swedish or having a currently Swedish-dominant daily language exposure did not influence Arabic comprehension. Similarly, attending MTI classes and performing joint reading activities at home did not influence Arabic comprehension as measured with the CLT at the age of 4–7. Hence, the regression model shows that older children score higher on Arabic vocabulary (CLT) comprehension than younger children and that children whose parents speak with them only/mostly in Arabic score higher that children whose parents do not speak with them only/mostly in Arabic.

In order to examine the combined effect of background factors on Arabic production, a multiple regression was conducted. Arabic production was the dependent variable and age (months), AoO of Swedish, parents’ L1 language with child, joint reading in Arabic, attending MTI, daily language exposure in Swedish and SES (binary) were the independent variables. Table 6.12 presents the regression model for Arabic production.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>16.45</td>
<td>5.87</td>
<td>.007**</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.19</td>
<td>.086</td>
<td>.21</td>
<td>.035*</td>
</tr>
<tr>
<td>AoO of Swedish</td>
<td>3.31</td>
<td>2.28</td>
<td>.14</td>
<td>.152</td>
</tr>
<tr>
<td>Parents’ L1 language with child</td>
<td>6.02</td>
<td>2.88</td>
<td>.20</td>
<td>.040*</td>
</tr>
<tr>
<td>Joint reading in Arabic</td>
<td>5.81</td>
<td>2.41</td>
<td>.23</td>
<td>.019*</td>
</tr>
<tr>
<td>MTI attendance</td>
<td>4.16</td>
<td>2.50</td>
<td>.16</td>
<td>.101</td>
</tr>
<tr>
<td>Daily exposure to Swedish</td>
<td>-8.32</td>
<td>2.55</td>
<td>-.32</td>
<td>.002**</td>
</tr>
<tr>
<td>SES (binary)</td>
<td>-2.48</td>
<td>2.25</td>
<td>-.10</td>
<td>.273</td>
</tr>
</tbody>
</table>

R² (adjusted) .484

Note. * = p < .05, ** = p < .01, *** = p < .001. F(7,66) = 10.78, p < .001.

The model explains 48.4% of the variance in the children’s Arabic production scores. Similarly to Arabic comprehension, age and parents’ L1 language use to the child were significant predictors for Arabic production. Joint reading activities between the parents and the child also had a significant effect on Arabic production scores. While having a dominant daily Swedish language exposure did not affect Arabic comprehension, it did significantly predict Arabic production scores negatively. Notice that AoO of Swedish, SES and MTI were not significant predictors for Arabic production as measured with the CLT at the age of 4–7. The regression model thus shows that for Arabic vocabulary (CLT) production, older children score higher than younger children, children whose parents speak with them only/mostly in Arabic score higher than children whose parents do not, children whose parents often read with
them in Arabic score higher than children whose parents do not. Children who are exposed to Swedish the majority of their day score lower on Arabic vocabulary production than children who are exposed to more Arabic daily.

6.3.3.6.2 Multiple regression models for Swedish vocabulary scores

In order to examine the combined effect of background factors on Swedish comprehension, a multiple regression was conducted. Swedish comprehension was the dependent variable and age (months), AoO of Swedish, joint reading in Swedish, and daily language exposure in Arabic were the independent variables. Table 6.13 presents the regression model for Swedish comprehension.

Table 6.13. Linear regression model of the combined effect on Swedish (CLT) comprehension scores.

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>16.49</td>
<td>4.28</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>Age</td>
<td>.46</td>
<td>.06</td>
<td>.62</td>
</tr>
<tr>
<td>AoO of Swedish</td>
<td>6.75</td>
<td>1.67</td>
<td>.32</td>
</tr>
<tr>
<td>Joint reading in Swedish</td>
<td>1.59</td>
<td>1.68</td>
<td>.08</td>
</tr>
<tr>
<td>Daily exposure to Arabic</td>
<td>-4.96</td>
<td>1.76</td>
<td>-.22</td>
</tr>
</tbody>
</table>

The model explains 55.6% of the variance in the children’s Swedish comprehension scores as measured by the CLT between ages 4–7. Age and AoO of Swedish were significant predictors for Swedish comprehension scores. Likewise, being exposed to Arabic the majority of the day was a significant predictor. Parent and child joint reading activities in Swedish was not a significant predictor of Swedish comprehension scores in the regression analyses. Thus concerning Swedish vocabulary (CLT) comprehension, older children score higher than younger children, children with an early AoO of Swedish score higher than children with a late AoO of Swedish, and children who are not exposed the majority of their day to Arabic score higher than children who are exposed to more Arabic daily.

Finally, in order to examine the combined effect of background factors on Swedish production, a multiple regression was conducted. Swedish production was the dependent variable and age (months), AoO of Swedish, parents’ L1 (Arabic) language with child, joint reading in Swedish, and daily language exposure in Arabic were the independent variables. Table 6.14 presents the regression model of the Swedish production scores.
<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.41</td>
<td>5.09</td>
<td>.506</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.44</td>
<td>.07</td>
<td>.54</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>AoO of Swedish</td>
<td>7.36</td>
<td>1.97</td>
<td>.32</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>Parents’ L1 language with child</td>
<td>2.22</td>
<td>2.28</td>
<td>.08</td>
<td>.332</td>
</tr>
<tr>
<td>Joint reading in Swedish</td>
<td>1.57</td>
<td>1.99</td>
<td>.07</td>
<td>.433</td>
</tr>
<tr>
<td>Daily exposure to Arabic</td>
<td>-6.00</td>
<td>2.10</td>
<td>-.24</td>
<td>.006**</td>
</tr>
<tr>
<td>R² (adjusted)</td>
<td>.483</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * = p < .05, ** = p < .01, *** = p < .001. F(5,77) = 16.31, p < .001.

The model explains 48.3% of the variance in Swedish production scores, showing that age and early AoO of Swedish were significant predictors. Additionally, having an Arabic dominant daily language exposure was also a significant predictor of Swedish production scores. When included in the regression model, parents’ L1 language to the child and joint parent-and-child reading in Swedish showed not to be significant predictors for Swedish production as measured by the CLT between the ages 4–7. Hence, for Swedish vocabulary (CLT) production, older children score higher than younger children, children who have an early AoO of Swedish score higher that children who have a late AoO of Swedish, and children who do not have the majority of their daily language exposure to Arabic score higher than children who have the majority of their daily language exposure to Arabic.

6.3.3.7 Summary

First, the individual effects of various background factors (AoO of Swedish, SES, different measures of language use in and outside the home, and current daily language exposure) were analysed (for the whole sample as a group) for Arabic comprehension, Arabic production, Swedish comprehension and Swedish production scores. Then, significant factors were included in multiple regression models to calculate the joint effect that the factors had on the dependent variables.

Several background factors were not significant predictors when included with other factors in the regression models. For both Arabic comprehension and production, age and parents’ L1 language use with the child were significant predictors. Additionally, for Arabic production, joint parent-and-child reading in Arabic as well as having a high daily language exposure to Swedish were also significant predictors. For Swedish comprehension and production, age, AoO of Swedish and having a high daily language exposure to Arabic were significant predictors.
6.4 Discussion

This section contains a discussion regarding the findings in the cross-sectional study in Sweden in relation to previous studies. Concerning vocabulary scores and age, the research questions addressed in the present chapter were:

- Is there a difference between the vocabulary scores in Arabic and Swedish?
- How does vocabulary develop with age in Arabic and Swedish?
- Is there a difference between the comprehension scores and the production scores in each language?

Concerning vocabulary scores and background factors, the research questions addressed were:

- Is there a relationship between the age of onset of Swedish and Arabic and Swedish vocabulary comprehension and production?
- How does language use in the home affect Arabic and Swedish comprehension and production?
- How does language use outside the home affect Arabic comprehension and production?
- How do age and background factors combined affect Arabic and Swedish vocabulary?

Unsurprisingly, at group level, the children scored significantly higher in comprehension than in production (in line with other studies that have also used the CLT, e.g. Bohnacker et al., 2016; Haman et al., 2017; Lindgren, 2018; Öztekin, 2019). For both languages, there was more variation in the scores for production than for comprehension. A wide range of the scores was found for all age groups (except for the Arabic comprehension scores where older children consistently scored higher than the younger ones). The results on the CLT did not show any floor effects or ceiling effects in any language, however, some of the 5-, 6-, and 7-year-olds scored near ceiling on the comprehension tasks.

6.4.1 Age

There was no significant difference between the Arabic and Swedish scores in neither comprehension nor production for any age group. The only exception was in the youngest age group (4-year-olds) where the children scored higher on Arabic comprehension than on Swedish comprehension. A possible explanation can be related to the young age of the young children and the age of onset of their respective languages. At age 4, all the children in the present study attended preschool and had not had any formal literary instruction in the majority language. Furthermore, some children had not had substantial input
in Swedish prior to the testing occasion, in contrast to Arabic which was heard from birth.

Despite the generally large individual variations in the scores (wide score range), the effect of age was strong and positive for both receptive and expressive vocabulary in both languages; however, it was stronger in Swedish than in Arabic. The mean scores for the 4-year-olds in Swedish comprehension (M = 36.1) and production (M = 22.3) were lower than those for Arabic (comprehension M = 41.6, production M = 25.6). By age 7, the Swedish mean scores were slightly higher than the Arabic ones (Swedish: comprehension M = 53.3, production M = 39.5; Arabic: comprehension M = 52.4, production M = 37.1). When age (in months) was put in the regression models with other variables, it was a significant predictor of both Arabic and Swedish comprehension and production scores.

Keeping in mind that this is a cross-sectional study rather than a longitudinal one, the study’s results go in line with findings from the literature. The fact that the majority language (in this case Swedish) increased more strongly with age echoes findings of international studies (e.g. Cobo-Lewis et al., 2002a, 2002b; Armon-Lotem et al., 2011; Gathercole et al., 2013; Hoff et al., 2014; Prevo et al., 2014; Montanari et al., 2019) and studies from the Swedish context (Ganuza & Hedman, 2019; Lindgren, 2018; Öztekin, 2019; Lindgren & Bohnacker, 2020).

However, not all studies have found an increase in the home language vocabulary (e.g. Verhoeven & Boeschoten, 1986; Prevo et al., 2014). In Sweden, for example, Bohnacker et al. (2016) studied the expressive vocabulary (using the CLT) of 40 Turkish-Swedish-speaking children and 38 German-Swedish-speaking children (all aged 4 –6) and found no correlation between the children’s age in months and the production scores in the minority language (Turkish or German). Furthermore, Lindgren (2018), who studied the expressive vocabulary of 72 Swedish-speaking monolingual, 48 Turkish-Swedish-speaking bilinguals, and 46 German-Swedish-speaking bilinguals (all aged 4 –6), found that there was an increase in scores with age for the majority language for all groups, but no increase in the productive vocabulary in the bilingual children’s minority language.169

Possible explanations for why the Arabic-Swedish-speaking children’s vocabulary scores increase with age in the home language in addition to the majority language could be related to several factors. The bilingual families in the present study speak Arabic, a minority language that is not the societal language, but can be considered to have high ‘ethnolinguistic vitality’ (Hardwood, Giles & Bourhis, 1994).170

169 The children who participated in Lindgren (2018) also include the children who participated in Bohnacker et al. (2016).
170 Hardwood, Giles, & Bourhis (1994, pp. 167–168): “The vitality of an ethnolinguistic group was defined as ‘that which makes a group likely to behave as a distinctive and collective entity within the intergroup setting’ (Giles, Bourhis & Taylor, 1977, p. 308). It is proposed that the
The Arabic-speaking population in Sweden is quite large (recall Section 1.2) which gives Arabic-speaking children opportunities to practice the home language outside of home, especially if the family lives in one of Sweden’s largest cities or their suburbs (where a lot of the data was gathered from for the present study). About one fourth of the children were reported to hear Arabic from friends/playmates and/or (pre)school staff. Associations and congregations where Arabic speakers gather are found in abundance, especially in Sweden’s largest cities (recall that several children were recruited via such organizations). Although there might be great variation as to how much Arabic is spoken and heard outside of home, nonetheless, there are opportunities for Arabic-Swedish-speaking children to be exposed to the home language outside of home. Recall also that 78% of the children were reported to hear Arabic from extended family members. So, even if children do not take part in activities where other Arabic speakers are present, many of the children are exposed to Arabic from other sources. In general, there seems to be a wish in Arabic-speaking families, living in societies where Arabic is not the majority language, to maintain and develop their mother tongue. In the parental questionnaire, the majority of the parents (81%) expressed that it was important for them that their child learned both Arabic and Swedish, while 12% of the parents expressed that learning Arabic was the most important. In the longitudinal study where the author had one-on-one contact with the parents during the informal interview (see Section 8.3.5), parents expressed their wish for their child to learn Arabic, foremost, in order to be able to speak with non-Swedish-speaking relatives, to communicate with native speakers when travelling back to the home country, and to understand/participate in religious ceremonies.

As previously mentioned, there were substantial individual differences between the children in both vocabulary measures of Arabic and Swedish, more clearly visible for expressive vocabulary where variation in the scores (in both languages) did not decrease with age. This suggests that vocabulary development in bilingual children is strongly affected by variables other than (chronological) age. The effect of different background factors found in the present study will be discussed below in relation to previous studies.

more vitality an ethnolinguistic group has, the more likely that it will survive and thrive as a collective entity in the intergroup context. Conversely, it was suggested that ethnolinguistic groups that have little or no vitality would eventually cease to exist as distinctive linguistic groups within the intergroup setting.”

171 These opinions were expressed when interviewing the parents of 10 children out of the 100 children participating in the study. These statements may and may not reflect the opinions of the remaining families.
6.4.2 Age of onset

The first factor that was investigated was the effect of age of onset (AoO) of Swedish. A comparison was made between the vocabulary scores of simultaneous vs. sequential bilinguals where AoO of Swedish at 3 was the cut-off point. Results showed that AoO of Swedish was a significant predictor of Swedish comprehension and production scores and remained as such when it was included in the regression analysis model together with other individually significant variables.\textsuperscript{172}

Furthermore, because the waking time of a bilingual child is divided between two languages, this study also examined whether AoO of the majority language had any effect on the vocabulary of the home language. When the effect of AoO of Swedish was examined alone, it showed a significant negative effect on both Arabic comprehension and production scores. However, when AoO of Swedish was included in the regression model to examine its combined effect with other background factors, AoO of the majority language was not a significant predictor. Hence, other language input factors have a stronger effect on Arabic vocabulary than how early bilingualism started (i.e. how early the child’s language input became divided between two languages).

While almost all children (98%) had an AoO of Arabic from birth, about half of the children (48%) started to have a regular exposure to Swedish before age 3. This might explain why the score range for Swedish comprehension and production scores remained quite large, even in the older age groups. Bohnacker et al. (2020) studied a subsample of the children ($N = 76$, age 4–6) in the present study and compared their scores on the Swedish CLT to an equivalent number of Turkish-Swedish-speaking children with the same age range. They found that the Turkish-Swedish-speaking children, as a group, scored significantly higher in Swedish production than the Arabic-Swedish-speaking children. The authors speculated that this result was related to the difference in the two groups’ place of birth and length of residency in Sweden (more Turkish-Swedish-speaking children were born in Sweden), in addition to the children’s AoO of Swedish (only 50% of the Arabic-Swedish-speaking children were exposed to Swedish before the age of 3 in comparison to 84% of the Turkish-Swedish-speaking children). In Bohnacker et al. (2021), the whole sample in the present study ($N = 100, 4–7;1$) was compared to an equivalent Turkish-Swedish-speaking sample ($N = 102, 4–8;1$) regarding the two bilingual groups’ Swedish vocabulary in relation to length of exposure.

\textsuperscript{172} Öberg (2020), who studied 99 of children in the present study, calculated the effect of Length of Exposure (LoE) of Swedish on both comprehension and production and found LoE to be an even stronger predictor of Swedish vocabulary than age (in months). In the regression model of the present study, age (in months) had a greater beta coefficient ($\beta$) than AoO of Swedish (binary split) for both Swedish comprehension and production, meaning that age (in months) had a stronger effect than AoO of Swedish on Swedish vocabulary.
(LoE) of Swedish. Although the older children in the Turkish-Swedish-speaking sample showed less variation in vocabulary scores than the older children in the Arabic-Swedish-speaking sample, when two groups’ performance on Swedish comprehension and production was compared taking into consideration the children’s LoE of Swedish (in linear regression analyses), the difference between the two groups disappeared. Hence, the large variation in the Swedish vocabulary scores for the Arabic-Swedish-speaking sample was a clear effect of short exposure to Swedish.

6.4.3 Socio-economic status (SES)

Another background factor that has been often been investigated in international studies is the effect of socio-economic status (SES) on vocabulary development. In the present study, the parents’ mean level of education was used as a proxy for SES, where high-SES indicates a completed secondary education and at least some tertiary education, i.e. higher than 3 on the ISCED scale (UNESCO Institute for Statistics, 2012). SES proved to have no significant effect on the children’s vocabulary in either language. When examined separately, a significant effect of SES was only found for Arabic production (with small effect size). However, SES was not a significant predictor of Arabic production when it was included in the regression model with other factors. The fact that children from a high-SES background did not score higher on L2 vocabulary than children from a low-SES background goes against many international studies (e.g. Leseman, 2000; Cobo-Lewis et al. 2002; Buac et al., 2014; Calvo & Bialystok, 2014). High SES is often associated with better opportunities for education and language input. However, from a Swedish perspective, SES seems to play a smaller role in children’s vocabulary development in either language of a bilingual child, at least for the age range that was investigated. Bohnacker et al. (2021) and Öztekin (2019), who used the same proxy for SES as the present study, also examined the effect of SES on Arabic-Swedish-speaking children and Turkish-Swedish-speaking children’s vocabulary in both languages and found no effect on either vocabulary measure in any of the children’s languages.

However, SES can be operationalized in several other ways (e.g. parents’ occupation, mother’s education, family income, and area of residency). In Sweden, child day-care is affordable for all families, regardless of family income. Children have the opportunity to enter preschool starting from a very young age (1 year). This may reduce potential differences in language input from home, at least concerning language input in the majority language. This could be a possible reason why no SES effects were found in either language

---

173 Even when SES was cut in three values (low-mid-high) or alternatively treated as a continuous variable (0–8 on the ISCED 2011 scale), no significant effect of SES was found in any language or vocabulary measure.
of the Arabic-Swedish-speaking children, nor the Turkish-Swedish-speaking children since potential effects of SES may even out in preschool where children have similar language input opportunities (in the majority language).

Another reason why no effects of SES were found in the present study may be related to the test material used to measure vocabulary. The CLT measures part of the children’s lexical breadth by examining both comprehension and production (of nouns and verbs). Other studies that have found effects of SES may have tested either one of these vocabulary measures and used other vocabulary tests (or parental checklists on toddlers). Potential effects of SES might be visible using other lexical tests than the one used in the present study.

Furthermore, effects of SES might be more visible when children become older. For example, a recent statistical report from the National Agency for Education (2021b) shows that, as a group, 15–16-year-old children where at least one parent has a tertiary degree, have better academic achievements than children where neither parent has a tertiary education. The children in the present study are quite young (4–7) and are merely tested on vocabulary. Effects of SES might not be noticeable when children are this age.

Another reason why SES did not affect the vocabulary scores in the majority language could have to do with the parents’ language use with the child. High SES is usually associated with more use of the majority language at home (Prevoo et al., 2014; Buac et al., 2014), however, in the present study, the vast majority of the households (79%) reported that both parents spoke mainly Arabic with the child. Several studies that have examined young children (as young as those in the present study or even younger) have argued that language input factors, such as parental speech and home language input, may influence language development to a higher degree than SES (e.g. Bohnacker et al., 2016; Bohnacker et al., 2021; Hart & Risley, 1995; Hoff, 2003; Calvo & Bialystok, 2014; Meir & Armon-Lotem, 2017; Rowe, 2012). The effect of these factors will be discussed below.

6.4.4 Input

Parents’ language use with the child was used as a variable in the present study since it is an important source for daily language input. The results showed that when parents spoke mostly Arabic with their child, this had a significant and positive effect on the child’s home language vocabulary in both comprehension and production. When parents’ L1 language use with the child was included in the multivariate model, it was a significant predictor of both Arabic vocabulary measures. Bohnacker et al. (2016), Bohnacker et al. (2021) and Öztekin (2019) also found a significant positive effect of parents’ language

---

174 No speech samples from the parents were collected in the present study, hence no qualitative analyses of the parents’ language use was made (such as studying the use of sophisticated language, rare words and decontextualized language, e.g. Rowe, 2012).
use for Turkish-Swedish-speaking children, where children whose both parents spoke mostly Turkish with the child, scored significantly higher on Turkish CLT than children whose parents spoke another language (or language combinations) with the child. Similar results were found in Bohnacker et al. (2016) for German-Swedish-speaking children where parents’ language use with the child significantly predicted CLT scores on the home language.

Additionally, results showed that parents’ use of Arabic at home did not negatively affect the children’s Swedish vocabulary. Recall that the variable ‘parents’ L1 language use with the child’ showed a significant negative effect on the children’s Swedish production scores when analysed alone, however, when included in the multivariate analyses, it was not a significant predictor. This means that other factors affect the development of the majority language’s vocabulary production more strongly than parental language input at home. This result is in line with Montanari et al. (2019) who found that the use of the L1 (Turkish) at home benefited the children’s L1 vocabulary without negatively impacting the vocabulary of their L2 (German).

In the present study, factors that were related to the child’s language output or that were documented without detailed precision (e.g. via yes/no questions), were not included in the multivariate models. The reason for this is related to the difficulty in measuring the actual effect of these variables since the frequency (how often), amount of time (how long), as well as number of interlocutors (how many friends/playmates/relatives) is not known. This makes these variables (in the way information was gathered in the present study) not sufficiently reliable sources of information regarding language input. For comparison, ‘parental language use’ was documented for each parent on five-point scales (with the option to add an ‘other’ option), while hearing the home language from friends/playmates was asked via a multiple-choice question where parents were asked to mark whether their child heard the home language from one or several sources, namely siblings, friends, extended family, book reading, TV/computer.

Still, these factors give us an idea of the use of Arabic outside the home. Out of the variables hearing the home language from extended family members, friends/playmates or staff members, hearing Arabic from friends/playmates was the only one with a measureable significant positive effect on Arabic production. Bohnacker et al., (2016) also found a significant interaction effect between parents’ language use to the child and hearing the home language from friends for Turkish-Swedish-speaking and German-Swedish-speaking children: for children who did not receive home language input from both parents, having additional home language input from friends, boosted the children’s vocabulary production scores in the home language.

However, not finding a statistically significant effect does not mean that hearing the home language from other sources (e.g. extended family members or school staff) does not positively affect vocabulary development. Such effects might be more visible if data included more information regarding the
language input frequency and intensity. Furthermore, the uneven distribution in the age groups for both variables, where the majority of the younger children had Arabic-speaking school staff members, and the older children had more contact with Arabic-speaking extended family members might have hidden the effect of such language input.

The parents’ language use with the child mirrored that of the children’s language use with the parents. However, input in a certain language invites more use of the language (Pearson, 2007). In order to avoid collinearity, language output from the child was investigated separately and was not included in the multivariate models. Children who spoke mostly Arabic with their parents scored higher on Arabic comprehension and production, but lower on Swedish comprehension and production than children who did not. Likewise, children’s Arabic language use with siblings had significant positive effect on both Arabic vocabulary measures and negative effects on the Swedish vocabulary measures. The same held true for children’s use of Swedish with siblings, with positive effects on both Swedish vocabulary measures and negative effects on Arabic production but no effect on Arabic comprehension.

However, having an older sibling did not prove to be a significant advantage for the home language nor for the majority language. This finding is not line with Bridges & Hoff (2014) who found that toddlers who had an older sibling had more advanced L2 (English) vocabulary development than toddlers who did not have an older sibling. Furthermore, those children who had an older sibling and were mostly spoken to in L2 by their siblings, scored higher than those who heard both L1 and L2. Sorenson Duncan and Paradis (2020) found that the relative quantity of L2 input by older siblings was significantly positively related to children’s L2 receptive vocabulary. In the present study, no advantage of having an older sibling was found for any language or vocabulary measure. However, it is important to keep in mind the diversity of the children’s backgrounds in the present study. Some families had only been in Sweden for a relatively short time prior to the testing session, therefore having an older sibling would not necessarily mean that the older sibling knew the majority language (or had been exposed to it) any more than the younger child. Hence, the siblings’ (and the family’s) short length of exposure to the majority language might have contributed to upkeep of the home language rather than bringing in more L2 input into the home (see Paradis et al., 2020). Future analyses could measure whether language input (mostly Arabic vs. mostly Swedish) from older siblings has an effect on vocabulary scores.

Other language input possibilities occurring at home include parent-child joint book reading. When analysing the effect of this variable alone, joint book

---

175 In 79% of the households, both parents spoke mainly Arabic with the child and in 68% of the households the child spoke mainly Arabic with the parents, however, only 39% of the children chose to speak mostly Arabic with their sibling (while 36% spoke both Arabic and Swedish and 18% spoke mostly Swedish).
reading in Arabic and joint book reading in Swedish showed significant positive effects on the languages’ respective vocabulary measures. However, recall that the distribution of the children for this variable was not even in the age groups; older children were reported to read more with their parents than younger children. When the independent variable was included in the regression models, joint book reading only showed a positive effect for Arabic vocabulary production but not for Arabic comprehension and Swedish comprehension and production (as measured with the CLT).

Several previous studies have shown a positive effect of joint book reading on receptive vocabulary (Frijters et al., 2000; Sénéchal & LeFevre, 2002, Deckner et al., 2006; Demir-Lira et al., 2018; Breitfeld et al., 2021; Prevoo et al., 2014) while a few others have not (Evans et al., 2000; Sim, 2012). The majority of these studies included monolingual children or children who were tested in the majority language only. For bilingual children, Prevoo et al. (2014) found positive effects on the expressive vocabulary of L2 and receptive vocabulary of L1, while Quiroz et al. (2010) found positive effects on the expressive vocabulary of both languages of bilingual children in their respective languages.

In the present study, the effect of joint book reading in the majority language was not a significant predictor of Swedish vocabulary comprehension and production when it was included in the multifactorial model. The reason could be related to the already high number of hours that a child is exposed to the majority language during the day. Therefore, reading an extra half an hour or so at home might not show remarkable differences in the children’s Swedish vocabulary as measured by the CLT.

The effect of joint book reading was, however, a significant positive predictor of vocabulary production in the home language, but not for comprehension. One possible explanation could be related to the diglossic nature of the Arabic language and the CLT comprehension task used in the present study. As previously mentioned in Chapter 1, Arabic is usually read and written in MSA (and not in the colloquial variety). Hence, joint book reading at home probably also occurs in MSA (or a with a mixture of both colloquial and MSA), while the CLT comprehension task measures the children’s colloquial Arabic knowledge of the test items.

The effect of attending mother tongue instruction (MTI) classes was also measured in the present study since Sweden is one of few countries that offer state-funded MTI for bilingual children. When considered as an individual variable, children who attended MTI performed better on Arabic comprehension and production than children who did not. However, when included in the regression model with other variables, MTI was not a significant predictor of both Arabic vocabulary measures.

One explanation could be related to the above-mentioned diglossic nature of Arabic and the difference between spoken and the written Arabic. During
MTI classes, children are taught foremost in MSA, which is not what is assessed in the present study. Hence, measuring the effect of attending MTI classes with the CLT (varietal) comprehension task might not be the most accurate choice of measurement concerning the advantages of attending MTI. On the other hand, for the CLT production task, children may have provided the target word in MSA (which would have been scored as correct). However, the whole session was in spoken colloquial Arabic, not MSA, so answering in MSA might not be the first thing to come to mind.

Another explanation as to why MTI was not a significant predictor of the Arabic vocabulary measures could be related to the uneven distribution in the age groups. Recall that more older children (6-year-olds and 7-year-olds) attended MTI than the younger children. This could explain why the significance of the MTI attendance disappeared when age (as a continuous variable) was also taken into account in the regression analyses.

Yet another explanation could be related to the age of the children participating in the study. MTI classes organized by the municipality usually comprise weekly 40–60 minute sessions. (Some children additionally attend MTI arranged by private initiatives that usually last for the same amount of time.) Hence, the effect of accumulated MTI attendance might not be as clearly visible in young children, but become more visible when children become older. Gauna & Hedman (2019) did find measurable effects of MTI attendance on the vocabulary in Somali-Swedish-speaking children. However, the children in their study (age 6–12) were slightly older than the children in the present study (age 4–7). Additionally, they were tested using other vocabulary and language tasks. Bohnacker et al. (2021) also investigated the effect of MTI attendance on the Arabic-Swedish-speaking children in the present study and on same aged Turkish-Swedish-speaking children. Effects of MTI were found for the Arabic-Swedish-speaking children’s vocabulary but not for the Turkish-Swedish-speaking children’s vocabulary. However, upon closer inspection of the two bilingual groups regarding hours of MTI attendance, the Turkish-Swedish-speaking children were reported to attend MTI for half the amount of time than the Arabic-Swedish-speaking children (Arabic: Mean = 1.9 hours/week; Turkish: Mean = 0.9 hours/week). Thus, the number of hours that children attend MTI might have an impact on whether the children benefit from these classes or not.

A final language input variable that was taken into consideration in the present study is the child’s current daily language exposure. The parents were asked to estimate their child’s daily language exposure (on a 7-point scale) which was later classified into three categories ‘mostly Arabic’, ‘even exposure’, and ‘mostly Swedish’. For both Arabic and Swedish, children who received more input in one language scored significantly higher than children

---

176 Bohnacker et al. (2021) studied the effect of MTI as a separate factor, and not as part of a multivariate analysis.
who received language input mostly in the other language. This is in line with several studies that emphasize the importance of language input (e.g. Hoff et al., 2012; Pearson et al., 1997; Sorenson Duncan & Paradis, 2020; Thordardottir, 2011; Unsworth, 2016). In the present study, for both the home language and the majority language, daily language exposure had a larger impact on vocabulary production than on comprehension, which in line with previous studies (e.g. Thordardottir, 2011; Buac et al., 2014; Öztekin, 2019).

For Swedish, having the majority of one’s daily language exposure in Arabic significantly negatively affected the child’s Swedish vocabulary comprehension and production scores. However, in both regression models, the beta coefficient (β) for ‘daily language exposure in Arabic’ was smaller than that for age and AoO of Swedish (the only other two significant factors in the model), meaning that the child’s age and AoO of Swedish were stronger predictors for Swedish vocabulary than having the majority of the daily language exposure in Arabic. For Arabic, having one’s daily language exposure mostly in Swedish was not a negative predictor of Arabic comprehension but was a negative predictor of Arabic production. Although a child might have the majority of her/his daily language exposure in Swedish, other factors, such as age, parents’ L1 language input and joint book reading in Arabic, positively affect the child’s development of the home language vocabulary production.

6.4.5 Arabic varieties

Lastly, the issue of the various Arabic spoken varieties by the children is discussed. It was decided early on in the data collection to limit the Arabic varieties to those spoken in the Levant and Iraq since these are the varieties that are primarily used by Arabic speakers in Sweden. Another important reason to limit the Arabic varieties was to match the varieties spoken by the experimenter and the child, where the main aim was to avoid any experimenter-child communication difficulties.

The author, a native Lebanese-Arabic speaker, tested the vast majority (N = 91) of the children in the study. Two other Lebanese-Arabic-speaking research assistants tested 3 other children, and an Iraqi-Arabic-speaking research assistant tested 6 children (out of the 17 children) who spoke Iraqi-Arabic. All of the experimenters were informed about the Arabic variety that the child spoke ahead of the session and used the corresponding scoring sheets with the appropriate instructions. Hence, all the children speaking a certain variety were tested using the same test sheet and instructions, irrespective of the Arabic variety of the experimenter.177 Despite these precautions and preparations, it is possible that children who were tested by an experimenter who was a native speaker of the same Arabic variety as their own scored higher

177 Additionally, recall that for the CLT tasks, there are 2 practice items at the beginning to make sure that the child has understood the instructions for each task.
than children who were tested by an experimenter who spoke another variety. Below, examples from two groups of children will be discussed: children speaking Lebanese Arabic (i.e. the colloquial Arabic variety of the author who tested these children) and children speaking the Iraqi dialect(s) (i.e. the Arabic variety that differs the most than from of the author, out of the Arabic varieties that participated in the present study).

As previously mentioned, children who spoke Lebanese Arabic ($N = 9$) were tested by a Lebanese-Arabic-speaking experimenter, however, these children did not necessarily perform better than their same aged peers. In fact, 6 of these children (BiAra4-02, BiAra5-25, BiAra6-01, BiAra6-19, BiAra7-12, BiAra7-13) scored below the mean score of their age group on both Arabic comprehension and Arabic production. These children had either an even daily language exposure, or the majority of their daily language exposure in Swedish. Furthermore, the children had both or one of their parents speak mostly Swedish with them. (Only one 6-year-old Lebanese-Arabic-speaking child was addressed in mainly Arabic by both parents.) However, all 6 children scored higher on Swedish comprehension and production than on Arabic comprehension and production, respectively.\footnote{The three children (3/9) who scored much higher than their age group mean were two 4-year-olds and one 5-year-old who were reported to hear the majority of their daily language exposure in Arabic and whose both parents spoke only/mostly in Arabic with them.} Hence, background factors related to the children’s daily language exposure and the parents’ language use with the child may provide a plausible explanation for the children’s low scores on Arabic vocabulary. The similarity in the variety spoken by the experimenter and the Lebanese-Arabic-speaking children did not advantage the children’s Arabic vocabulary scores.

Children who spoke Iraqi Arabic and who were tested by a native Iraqi-Arabic-speaking experimenter did not score higher than children who were tested by a non-Iraqi-speaking experimenter. When looking at the performance of the individual Iraqi-Arabic-speaking children, great variations in daily language exposure can explain the children’s scores. Furthermore, clear patterns of language input from the parents and an early AoO of Swedish can be seen.\footnote{Similar observations concerning the effect of daily language exposure and parental language use were made for the four Egyptian-Arabic-speaking children (BiAra4-14, BiAra5-16, BiAra5-24, BiAra7-02) in the study.}

In general, children who had the majority of their daily language exposure in Swedish, and/or had parents who spoke with them mostly in Swedish scored lower than their peers on Arabic comprehension and production, irrespective of which Arabic variety they or the experimenter spoke.
6.4.6 Concluding remarks

Summing up, the present study investigated the vocabulary of 100 Arabic-Swedish-speaking children (4–7) using the Cross-linguistic Lexical Task (CLT). Vocabulary comprehension and production in both of the children’s languages were explored in relation to age and a number of background factors. Despite a significant age development (at group level) for both Arabic and Swedish vocabulary (stronger for the majority language than the minority language), there were great individual variations in the scores, especially in vocabulary production. Background factors related to AoO and language use inside and outside the home were investigated individually to explain some of the variability in the results.

When all the factors that by themselves had a significant effect on the vocabulary scores were included together in regression models, only a limited number of independent variables were found to be significant predictors. For Arabic comprehension and production, age and parents’ L1 language use with the child were significant predictors. Additionally, for Arabic production, joint book reading in Arabic and daily language exposure in Swedish were also significant predictors. For Swedish, age, AoO of Swedish and daily language exposure in Arabic were significant predictors for both vocabulary comprehension and production.

For both Arabic and Swedish, the proportion of the variance that can be explained by the independent variables (adjusted $R^2$) was somewhat higher for comprehension than that for production, indicating that the models have a higher predictive power for receptive vocabulary than for expressive vocabulary. In other words, there are more variables that affect the vocabulary production of a bilingual child that have not been taken into account. However, in all four models, the $R^2$ (adjusted) were relatively high (55% Arabic comprehension, 48% Arabic production, 56% Swedish comprehension, 48% Swedish production) signifying that several of the important factors that affect the vocabulary development of Arabic-Swedish-speaking children were indeed investigated in the present study.
7. Cross-sectional study in Sweden: Narrative macrostructure

This chapter reports the results for narrative macrostructure from the cross-sectional study in Sweden, starting with the results for narrative comprehension (Section 7.1) followed by the results for narrative production (Section 7.2). The chapter ends with a discussion of the results (Section 7.3).

7.1 Narrative comprehension

This section starts with stating the specific research questions related to the children’s narrative comprehension development (7.1.1). Then, the data and the statistical methods used are described (7.1.2). After that, the results for narrative comprehension are presented in relation to age, language and story (7.1.3). Additionally, individual comprehension questions are analysed. The section ends with a summary of the narrative comprehension results.

7.1.1 Research questions

The following research questions concerning narrative comprehension are investigated:

- How does narrative comprehension develop with age in Arabic and Swedish?
- Do children perform similarly in their two languages with respect to narrative comprehension?
- Do children perform differently in the two narrative comprehension tasks (MAIN1 and MAIN2)?
- Do children perform differently on different types of narrative comprehension questions?

7.1.2 Data and statistical analyses

7.1.2.1 Data

The children were asked 10 comprehension questions after telling each MAIN narrative. If answered correctly, each question yielded 1 point. Out of the 100
children who participated in the cross-sectional study in Sweden, a few children did not answer the comprehension questions. Had all children answered all the comprehension questions, there would have been 2000 responses in each language (100 children x 2 stories x 10 comprehension questions). However, in both languages, a number of questions were not posed. In Arabic, one child (BiAra4-02) did not cooperate well during the Arabic testing session. As a result, the child did not tell MAIN1 and his MAIN2 was prematurely terminated and excluded from the analyses because the child only responded in the non-target language. In Swedish, two children (BiAra4-05 and BiAra4-19) were timid at the beginning of the testing session and consequently refused to tell MAIN1. One child (BiAra5-13) had limited concentration during the testing session and refused to tell both MAIN1 and MAIN2.

For some other children, the experimenter occasionally skipped or forgot to pose certain comprehension questions. The reason behind this was usually related to experimenter error or as a consequence of the child’s unclear speech/mumbling. For example, when the experimenter was not able to hear clearly what the child answered in the internal state term (IST) question, the follow-up rationale question could have been skipped. There were 12 such instances of missing data in each language.

Hence, for Arabic, a total of 20 questions were not carried out and 12 questions were not posed (missing data). This leads to a final total of 1968 comprehension responses in Arabic that were answered and could thus be scored. For Swedish, 40 questions were not carried out and 12 questions were not posed (missing data), meaning that a total of 1948 comprehension responses in Swedish were answered and could be scored. Table 7.1 contains a detailed overview of the number of questions in both languages and both narrative tasks. For the scoring of narrative comprehension, see Section 4.5.2.1.

Table 7.1. Narrative comprehension questions broken down from total potential responses to actual responses in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) in both languages.

<table>
<thead>
<tr>
<th></th>
<th>Arabic MAIN1</th>
<th>Arabic MAIN2</th>
<th>Swedish MAIN1</th>
<th>Swedish MAIN2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential responses</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Not carried out</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Not posed (missing data)</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Actually asked</td>
<td>986</td>
<td>982</td>
<td>966</td>
<td>982</td>
</tr>
</tbody>
</table>

7.1.2.2 Statistical analyses

The same analyses were performed on both languages (Arabic and Swedish) separately. MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) were analysed separately so that potential task and/or story effects could be explored. The level of significance was held constant at $p < .05$ (two-tailed) for all statistical analyses.
In the cases of missing data, i.e. when the child did not have the chance to hear the question and answer it, it would have been unfair not to award the child any points. Since the missing data did not exceed 2%, each missing score was substituted by the sample mean score as proposed by Widamann (2006). This means that, the missing value was substituted with the mean score for that particular question in that child’s age group, story and language. To illustrate, 4-year-old BiAra4-10 was not asked the follow-up question (‘Why do you think the cat is feeling hurt?’) in D3 in Arabic MAIN1 by mistake. Thus, BiAra4-10 was given the mean score (0.4 points) on question D3 for the 4-year-olds in MAIN1 (Arabic) as a substitute D3-score.

Independent samples t-tests were run to determine whether the children scored similarly on the different narratives of MAIN1 (Cat vs. Dog) and MAIN2 (Baby Birds vs. Baby Goats). In order to analyse whether there was a significant difference between the children’s scores in MAIN1 and MAIN2 and between the two languages, paired samples t-tests were run on the total scores (for all ages combined). The relationship between MAIN1 and MAIN2 comprehension scores was analysed with Pearson correlation (two-tailed). Independent samples t-tests were run to see whether the testing order (i.e. first testing in Arabic or Swedish) significantly affected the children’s results.

For both Arabic and Swedish, age effects for narrative comprehension scores were explored using one-way ANOVAs on the age groups. Post hoc analyses (with Bonferroni correction) were used to identify which age groups differed significantly from each other. Age was also operationalized as a continuous variable and simple regression analyses were run to analyse the relationship between the scores and age (in months). The regression analyses were visualized as regression lines on scatterplots showing the children’s individual scores in relation to age.

In addition to these statistical investigations of narrative comprehension total scores, response accuracies are reported descriptively for the individual comprehension questions and for certain types of comprehension questions, especially those targeting the understanding of goals.

This is done so that the children’s (prompted) comprehension of macro-structural components, such as goals, can be compared with the children’s production (or non-production) of these components in story telling.

7.1.3 Results

7.1.3.1 Overview of the results

This section reports the narrative comprehension scores for Arabic and Swedish. Table 7.2 displays the mean, standard deviation (SD) and ranges for both

---

180 For Arabic (MAIN1 and MAIN2 combined), there was a total of 0.61% missing data (12/1968) and for Swedish (MAIN1 and MAIN2 combined) there was a total of 0.62% missing data (12/1948).
languages in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) for all the children.

Table 7.2. Arabic and Swedish narrative comprehension scores for MAIN1 and MAIN2, all ages combined (4;0–7;11).

<table>
<thead>
<tr>
<th></th>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>6.37</td>
<td>6.66</td>
</tr>
<tr>
<td>SD</td>
<td>2.62</td>
<td>2.87</td>
</tr>
<tr>
<td>Range</td>
<td>1–10</td>
<td>0–10</td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.80</td>
<td>4.92</td>
</tr>
<tr>
<td>SD</td>
<td>2.36</td>
<td>3.13</td>
</tr>
<tr>
<td>Range</td>
<td>0–10</td>
<td>0–10</td>
</tr>
</tbody>
</table>

Note. Max narrative comprehension score = 10 points. BB = Baby Birds, BG = Baby Goats.

In order to see whether the children performed differently between the two stories within MAIN1 (Cat and Dog) and the two stories within MAIN2 (Baby Birds and Baby Goats), independent samples t-tests were run. No significant differences in performance were observed in either language between the two different stories in MAIN1 (Arabic: $t(95.48) = .740, p = .461$; Swedish: $t(94.30) = -.377, p = .707$) and the two different stories in MAIN2 (Arabic: $t(87.68) = .425, p = .672$; Swedish: $t(94.23) = 1.564, p = .121$). This means that, for both languages, the children scored equally well in Cat and Dog stories in MAIN1, as well as equally well in Baby Birds and Baby Goat stories in MAIN2. See Appendix A2.1 for a comparative table between the scores of MAIN1 (Cat and Dog) and MAIN2 (Baby Birds and Baby Goats) for all ages combined. Since there was no significant difference in the children’s performance between the stories of MAIN1 and MAIN2, the results are presented henceforth as MAIN1 and MAIN2, and not for the four stories separately.

Paired samples t-tests showed that there was no significant difference between the two languages’ MAIN1 scores ($t(95) = -.657, p = .513$) nor between their MAIN2 scores ($t(97) = -.524, p = .601$). As can be seen in Table 7.2, for both languages, the mean scores for MAIN1 were higher than for MAIN2. Paired samples t-tests confirm that the MAIN1 scores were indeed significantly higher than the MAIN2 scores (Arabic: $t(98) = 7.01, p < .001$; Swedish: $t(96) = 7.74, p < .001$). Hence, there was a task effect where children in both languages performed better in MAIN1 than in MAIN2. The standard deviation and ranges were quite similar for both languages and for both MAIN1 and MAIN2.

Pearson correlations (two-tailed) revealed that all narrative comprehension measures correlated positively with each other, but the correlation between MAIN1 and MAIN2 in the same language was stronger than the correlation with the other language (Arabic MAIN1 and Arabic MAIN2: $r = .604, N = 99, p < .001$, strong correlation; Swedish MAIN1 and Swedish MAIN2: $r = .751$, strong correlation).
When examining whether the order of the testing had any effect on the children’s narrative comprehension scores, independent samples t-tests revealed that for Arabic comprehension there was no effect for testing order neither for MAIN1 (\(t(97.00) = -1.43, p = .156\)) nor for MAIN2 (\(t(91.11) = -.588, p = .558\)). For Swedish comprehension, there was a significant effect for order of testing for both MAIN1 (\(t(94.40) = 3.11, p = .002, d = .632\), moderate effect size) and MAIN2 (\(t(96.73) = 2.01, p = .047, d = .404\), small effect size). In other words, children whose first testing session was in Swedish had considerably lower scores in Swedish narrative comprehension than children whose second testing was in Swedish (whilst their first testing was in Arabic), i.e. there was a training effect for children who were first tested in Arabic and then tested in Swedish.

Figure 7.1 gives an overview on the children’s narrative comprehension score for both languages.

![Figure 7.1](image-url)  
Figure 7.1. Bar graph over MAIN1 and MAIN2 comprehension for both Arabic and Swedish for all age groups. Error bars show ±1 SD.

As can be seen in Figure 7.1, for both languages, MAIN1 scores are higher than MAIN2 scores for all age groups. Also, there is an increase in mean

---

181 Similarly, Arabic MAIN1 and Swedish MAIN2 correlated (\(r = .368, N = 98, p < .001\), moderate correlation) and Arabic MAIN2 and Swedish MAIN1 correlated (\(r = .404, N = 96, p < .001\), moderate correlation).
scores across the age groups for both languages and for both MAIN1 and MAIN2.

Furthermore, paired samples t-tests showed that there was no significant difference between the two languages’ MAIN1 scores, nor between their MAIN2 scores at each age group level, not just for the sample as a whole, as mentioned above (4-year-olds, MAIN1: t(18) = .616, p = .545, MAIN2: t(20) = 1.145, p = .266; 5-year-olds, MAIN1: t(23) = -1.150, p = .262, MAIN2: t(23) = -1.841, p = .079; 6-year-olds, MAIN1: t(28) = -.282, p = .780, MAIN2: t(28) = .841, p = .407; 7-year-olds, MAIN1: t(23) = -.275, p = .786, MAIN2: t(23) = -1.048, p = .305).

In the following two sections, Arabic narrative comprehension (Section 7.1.3.2) and Swedish narrative comprehension (Section 7.1.3.3) are analysed in detail in relation to age. Individual cases of children who scored remarkably lower than their age group are discussed further in relation to certain background characteristics.

### 7.1.3.2 Arabic narrative comprehension

In Table 7.3, the mean, standard deviation (SD) and ranges for Arabic narrative comprehension scores in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) are presented for each age group.

**Table 7.3. Arabic narrative comprehension scores for MAIN1 and MAIN2 by age groups.**

<table>
<thead>
<tr>
<th></th>
<th>4-year-olds (N = 21)</th>
<th>5-year-olds (N = 25)</th>
<th>6-year-olds (N = 29)</th>
<th>7-year-olds (N = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.6</td>
<td>5.6</td>
<td>7.2</td>
<td>7.8</td>
</tr>
<tr>
<td>SD</td>
<td>2.5</td>
<td>2.2</td>
<td>2.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Range</td>
<td>1–10</td>
<td>2–10</td>
<td>1–10</td>
<td>3–10</td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.9</td>
<td>3.7</td>
<td>5.5</td>
<td>6.7</td>
</tr>
<tr>
<td>SD</td>
<td>1.7</td>
<td>1.5</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Range</td>
<td>0–7</td>
<td>1–6</td>
<td>2–10</td>
<td>2–10</td>
</tr>
</tbody>
</table>

*Note. Max narrative comprehension score = 10 points. BB = Baby Birds, BG = Baby Goats.*

Looking at Table 7.3, for MAIN1, the SD stays almost the same throughout the age groups but decreases in the 7-year-old group. As for the ranges, the oldest age group has the narrowest score range. For MAIN2, the 4-year-olds and 5-year-olds had a lower SD than the older two age groups. When looking at the ranges, no child in the 4-year-old or 5-year-old groups scored at ceiling; the maximum score for the 4-year-old group was 7 points and the maximum score for the 5-year-old group was 6 points. No child in the older two age groups scored at floor level; the minimal score was 2 points in both age groups.

When comparing MAIN1 with MAIN2, the children have a smaller SD in the 4-year-old and 5-year-old groups in MAIN2 compared to MAIN1. For the
6-year-old and 7-year-old groups, the SD is quite similar. Furthermore, there is a great difference between the ranges of MAIN1 and MAIN2 for the 4-year-olds and the 5-year-olds, but barely any difference between the ranges of MAIN1 and MAIN2 for the 6-year-old group and the 7-year-old group.

As for the mean scores, there is an increase for both MAIN1 and MAIN2 across the ages. One-way ANOVA tests confirm a significant difference between the age groups (MAIN1: $F(3,95) = 9.14, p < .001, \eta^2 = .224$, large effect size; MAIN2: $F(3,95) = 19.16, p < .001, \eta^2 = .377$, large effect size). A comparison between the eta-squared ($\eta^2$) values suggests that the effects of age are stronger for MAIN2 than for MAIN1.

Post hoc analyses (with Bonferroni correction) were used to find out between which age groups there were significant differences. For MAIN1, there was no significant difference between the 4-year-olds and the 5-year-olds ($p = .997$) but between the 4-year-olds and the older two age groups (4-year-olds vs. 6-year-olds: $p = .001$; 4-year-olds vs. 7-year-olds: $p < .001$). The 5-year-olds significantly differed from the 7-year-olds ($p = .008$) but not from the other two age groups (5-year-olds vs. 6-year-olds: $p = .074$). The 7-year-olds and the 6-year-olds did not differ significantly from each other ($p = 1.000$).

For MAIN2, there was a significant difference between the 4-year-olds and the oldest two age groups (4-year-olds vs. 5-year-olds: $p = .964$; 4-year-olds vs. 6-year-olds: $p < .001$; 4-year-olds vs. 7-year-olds: $p < .001$). Similarly, the 5-year-olds differed only from the two older groups (5-year-olds vs. 6-year-olds: $p = .005$; 5-year-olds vs. 7-year-olds: $p < .001$). There was no significant difference between the 6-year-olds and the 7-year-olds ($p = .123$).

As for the relationship between Arabic narrative comprehension and age as a continuous variable (in months), simple linear regression analyses were carried out to examine whether age (in months) predicted MAIN1 and MAIN2 narrative comprehension scores. For both MAIN1 and MAIN2, age was indeed a significant predictor (MAIN1: $F(1,97) = 28.99, p < .001, R^2 = .230$; MAIN2: $F(1,97) = 64.32, p < .001, R^2 = .399$). The $R$-squared value for MAIN2 ($R^2 = .399$) is nearly double that for MAIN1 ($R^2 = .230$), meaning that MAIN2 narrative comprehension scores increase much more strongly with age than MAIN1 narrative comprehension scores.

Figure 7.2 and Figure 7.3 illustrate the age development (in form of a regression line) on a scatterplot for Arabic MAIN1 narrative comprehension and Arabic MAIN2 narrative comprehension, respectively.
Figure 7.2. Scatterplot of Arabic MAIN1 (Cat/Dog) narrative comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

Figure 7.3. Scatterplot of Arabic MAIN2 (Baby Birds/Baby Goats) narrative comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

For Arabic MAIN1, about one third of the children \((N = 27)\) scored below 50% (below 5 points). No child scored at floor level (0 points). The majority of the children \((N = 59)\) scored between 50%–90% (between 5 and 9 points), while thirteen children scored at ceiling with 10 points. The ceiling scoring children belonged to all age groups (4-year-olds \(N = 1\); 5-year-olds \(N = 1\); 6-year-olds \(N = 6\); 7-year-olds \(N = 5\)).
A different pattern is observed for MAIN2 where about half of the children \((N = 51)\) scored below 50%. Forty-five children scored between 50%–90%, and only three children scored at ceiling. These three children were one 6-year-old and two 7-year-olds.

In what follows, children who scored the lowest in Arabic narrative comprehension in their age group will be discussed in relation to background factors, their performance in Swedish MAIN, and their Arabic and Swedish vocabulary (CLT) test.

In the 4-year-old group, the lowest scoring child in MAIN1 was BiAra4-05 with 1 point. The encounter with every child starts with doing the non-word repetition (NWR) task, which BiAra4-05 found very confusing and challenging. The child was shy at the beginning of the Arabic testing session but felt more comfortable as the session progressed. At the beginning of the Arabic testing session, it was quite hard for the experimenter to decipher what the child said due to hardly audible speech. BiAra4-05 scored 3 points on MAIN2. BiAra4-05 scored, however, well above the mean for his age group (Mean comprehension = 41.6, Mean production = 25.6) in both Arabic vocabulary measures (CLT comprehension 45/60, and production 38/60). In MAIN2, the lowest scoring child (0 points) was BiAra4-21 who also scored remarkably low on Arabic production CLT (12/60).\(^{182}\) Notably, this child did not score high in Swedish MAIN1 or MAIN2 either with 3 points and 1 point, respectively. BiAra4-21 was reported to have the majority of his daily language exposure in Swedish with additional daily exposure to English. In the 5-year-old group, BiAra5-06 scored the lowest amongst his peers (only 2 points in each of MAIN1 and MAIN2). This child also scored very low in the Arabic CLT tasks (comprehension 27/60, production 11/60), and was reported to have the majority of his daily language exposure in Swedish. Another low-scoring 5-year-old, BiAra5-21, scored low in MAIN2 (1 point) and was also reported to have the majority of his daily language exposure in Swedish. Unlike the previously mentioned children, BiAra5-21 scored quite high on Arabic vocabulary (comprehension 42/60, production 42/60). In the 6-year-old group, BiAra6-10 scored substantially lower than his peers with 1 point on MAIN1 and 3 points on MAIN2. BiAra6-10 was reported to have the majority of daily language exposure time in Swedish and scored high in both Swedish MAIN1 and MAIN2 with 10 points and 7 points, respectively.\(^{183}\) Table 7.4 contains the answers (and their English translation) provided by BiAra6-10 (age 6;7) in Arabic MAIN1 where the child scored almost at floor level. Notice the

\(^{182}\) BiAra4-21 scored also low on his Arabic MAIN1 with 3 points.

\(^{183}\) BiAra6-10 has already been mentioned in the vocabulary chapter since he scored substantially lower than his age group peers in Arabic vocabulary (CLT) (comprehension 32/60, production 16/60).
strong tendency towards code-switching to Swedish despite the exper-
imer’s efforts for an Arabic-only testing situation.¹⁸⁴

Table 7.4. Answers from a low scoring 6-year-old (BiAra6-10, age 6;7) on Arabic
MAIN1 (Cat) narrative comprehension questions.

<table>
<thead>
<tr>
<th>Question</th>
<th>Child’s answers</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.</td>
<td>‘a elfjäril [@s], fjäril [@s], elbarn [@s]</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>‘on the butterfly (Swedish), butterfly (Swedish), the child (Swedish)’</td>
<td></td>
</tr>
<tr>
<td>D2.</td>
<td>honi, kanat ki waqa’a fi elbuskar [@s] (.) w elbarn [@s] raša’at elkatt [@s] (.)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>w elboll [@s], elbisse raḥat ʿala elmay.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘here, she fell in the bushes (Swedish), (.) and the child (Swedish) threw the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cat (Swedish), (.) and the ball (Swedish), the cat went to the water’</td>
<td></td>
</tr>
<tr>
<td>D3.</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>D4.</td>
<td>&amp;ee elwalad ta yḡib elboll [@s] (child points at the ball)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘&amp;ee the boy to get the ball (Swedish)’</td>
<td></td>
</tr>
<tr>
<td>D5.</td>
<td>walad ḡab elboll [@s], aku stark [@s].</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>‘boy got the ball (Swedish), he was strong (Swedish)’</td>
<td></td>
</tr>
<tr>
<td>D6.</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>D7.</td>
<td>&amp;ee &lt;lite hungrig&gt; [@s]. (.) bi elʿarabi? (.) &amp;e fisk [@s].</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>‘&amp;ee &lt;a little hungry&gt; (Swedish), (.) in Arabic? (.) &amp;e fish (Swedish)’</td>
<td></td>
</tr>
<tr>
<td>D8.</td>
<td>&amp;e ken stark [@s].</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>‘he was strong (Swedish)’</td>
<td></td>
</tr>
<tr>
<td>D9.</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>D10.</td>
<td>(child nods) &lt;men annars&gt; [@s], ma ykunu asdiqā’. &amp;ee li yakun. (.) liʾan,</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&lt;dom älskar honom&gt; [@s].</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘(child nods), &lt;but otherwise&gt; (Swedish), they are not friends, yes, they will</td>
<td></td>
</tr>
<tr>
<td></td>
<td>become [friends], (.) because, &lt;they love him&gt; (Swedish)’</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Note. Max = 10 points.

In the 7-year-old group, BiAra7-16 scored low in MAIN1 with 3 points. The
child seemed to have a hard time understanding the questions related to internal
state terms (IST) and did not earn any point on any of the 3 IST questions
or their follow-up questions.¹⁸⁵ In MAIN2, BiAra7-19 scored only 3 points
although she had scored at ceiling with 10 points in Arabic MAIN1. The reason
could be related to the child’s misinterpretation of the character’s actions
in the MAIN2 pictures where the child thought that the baby goat was swim-
morning and that the mother was drinking water.

¹⁸⁴ BiAra6-10 also used the definite article el- ‘the’ in front of Swedish nouns, as well as paused
several times while answering, presumably to search for the appropriate Arabic word.
¹⁸⁵ BiAra7-16 is an Iraqi-speaking child and was tested by an Iraqi-speaking experimenter who
spoke the same Arabic variety. Still, this child seemed to have great difficulty in understanding
the intention behind the IST questions for both MAIN1 and MAIN2 in Arabic. BiAra7-16
scored 4 points on MAIN2, two out of the three questions probing for ISTs (and consequently
their follow-up questions) were answered incorrectly.
On the other end of the age range, a couple of 4-year-old and 5-year-old children scored high on MAIN1 (recall that ceiling was reached for MAIN1 even by a few children in the youngest age group). BiAra4-24 scored 10 points on MAIN1 and 4 points on MAIN2. The child was exposed mostly to Arabic (60%) throughout her day and scored higher than her peers also on Arabic vocabulary (CLT) comprehension (51/60) and production (40/60).

Another high scoring 4-year-old child was BiAra4-08 who scored 9 points on MAIN1 and 4 points on MAIN2. BiAra4-08 told her narration in MSA and answered the comprehension questions also in MSA. BiAra4-08’s answers on the comprehension questions for MAIN1 are found in Table 7.5.

| Table 7.5. Answers from a high scoring 4-year-old (BiAra4-08, age 4;11) on Arabic MAIN1 (Dog) narrative comprehension questions. |
|-----------------|------------------|
| Question        | Child’s answers                                                                 |
| D1.             | liʾannu [++] liʾannu yuridu ahḍ haḍāk alfār.                                      |
|                 | ‘because he wants to take that mouse’                                           |
| D2.             | istadam <bi sida> [: bi sudā’]. hek (child moves her body forward as if banging her head on the table.) |
|                 | ‘he bumped, a head ache, like this’                                            |
| D3.             | istadam <bi sida> [: bi sudā’]. hek (child moves her body forward as if banging her head on the table.) |
|                 | ‘he bumped, a head ache, like this’                                            |
| D4.             | liʾannu huwa yurid aḥḍa albalon.                                               |
|                 | ‘because he wants to take the balloon’                                         |
| D5.             | &eh, xx saʾid                                                              |
|                 | ’&eh, (unintelligible speech) happy’                                          |
| D6.             | liʾannahu aḥḍa albalon                                                        |
|                 | ‘because he took the balloon’                                                |
| D7.             | liʾannu ḥabb taʾmaha                                                           |
|                 | ‘because he loved it’s taste’                                                |
| D8.             | &el sawfu yaḏdab.                                                            |
|                 | ‘&el he will become angry’                                                    |
| D9.             | liʾannahu sayaʾkul almaqaqanik, unẓuri, aḥadha, unẓuri, ḥallas haḍa aliks. (CHI points at dog in pic6) |
|                 | ‘because he will eat the hotdog, look, he took them, look, he finished this bag’ |
| D10.            | (child nods) liʾanhum yaḥibun an yakunu latifīn.                              |
|                 | (child nods) ‘because they love to be nice’                                  |

Total: 9

Note. Max = 10 points.

186 Upon closer inspection of BiAra4-24’s MAIN2 (Baby Birds) narrative comprehension answers, the child misinterpreted why the mother bird was flying away (‘to buy something’) rather than specifying that she was bringing food for her children. Hence, the child scored 0 points on each of D1, D2 and D3. The child also misinterpreted the dog’s intention to rescue the baby birds (‘he was hungry, and wanted to eat the birds’) and as a result scored 0 points on each of D8, D9 and D10.
The child was reported to speak MSA at home and was exposed to the majority of her day (80%) to Arabic. BiAra4-08’s vocabulary (CLT) scores in Arabic were very high in both comprehension (52/60) and production (42/60).

In sum, for many of the low-scoring children on Arabic MAIN narrative comprehension, it seems that having a low vocabulary score on the Arabic CLTs and a low daily language exposure to Arabic were common features. Furthermore, several high scoring children also scored high on the Arabic vocabulary task as well as were reported to be exposed to Arabic the majority of their time.

7.1.3.3 Swedish narrative comprehension

Table 7.6 provides an overview (mean, standard deviation, range) of the Swedish narrative comprehension scores for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) divided per age group.

Table 7.6. Swedish narrative comprehension scores for MAIN1 and MAIN2 by age group.

<table>
<thead>
<tr>
<th></th>
<th>4-year-olds (N = 20/22)*</th>
<th>5-year-olds (N = 24)</th>
<th>6-year-olds (N = 29)</th>
<th>7-year-olds (N = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.4</td>
<td>6.4</td>
<td>7.4</td>
<td>8.0</td>
</tr>
<tr>
<td>SD</td>
<td>3.1</td>
<td>2.8</td>
<td>2.3</td>
<td>2.4</td>
</tr>
<tr>
<td>Range</td>
<td>0–10</td>
<td>2–10</td>
<td>2–10</td>
<td>2–10</td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.3</td>
<td>4.9</td>
<td>5.0</td>
<td>7.3</td>
</tr>
<tr>
<td>SD</td>
<td>2.2</td>
<td>2.9</td>
<td>2.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Range</td>
<td>0–8</td>
<td>0–10</td>
<td>1–10</td>
<td>2–10</td>
</tr>
</tbody>
</table>

*For the 4-year-olds, there were answers from 20 children for MAIN1 and from 22 children for MAIN2.

Looking at Table 7.6, for MAIN1, the SD was highest for the 4-year-olds, followed by the 5-year-olds and two oldest age groups. For the ranges, in all age groups ceiling scores were obtained, but only in the 4-year-old group did children score at floor level (0 points). For MAIN2, the SD for all age groups was quite similar. As for the ranges, the 4-year-olds did not score at ceiling (10 points) and the 6-year-olds and 7-year-olds did not score at floor level. Comparing MAIN1 and MAIN2, the SD and ranges seem quite similar across the age groups, except for the 4-year-olds who had a higher SD in MAIN1 and a narrower range in MAIN2.

187 The child was reported to attend a preschool where there were several Arabic-speaking staff members, and thus had the opportunity to hear her mother tongue to a large extent outside of her home. Unsurprisingly, BiAra4-08’s vocabulary (CLT) scores in Swedish were very low (comprehension 28/60, production 10/60).
As for the mean scores, an increase can be seen across the age groups for both MAIN1 and MAIN2. A significant difference was found between the age groups when one-way ANOVAs were run for both MAIN1 ($F(3,93) = 7.62, p < .001, \eta^2 = .197$, large effect size) and MAIN2 ($F(3,95) = 13.77, p < .001, \eta^2 = .303$, large effect size). Similarly to the Arabic narrative comprehension scores, comparing the eta-squared ($\eta^2$) values for the Swedish scores suggests that the effects of age are stronger for MAIN2 than for MAIN1.

For MAIN1, post hoc analyses (with Bonferroni correction) revealed that there was a significant difference between the 4-year-olds and the two oldest age groups (4-year-olds vs. 5-year-olds: $p = .105$; 4-year-olds vs. 6-year-olds: $p = .001$; 4-year-olds vs. 7-year-olds: $p < .001$). There was no significant difference between any other age group (5-year-olds vs. 6-year-olds: $p = .959$; 5-year-olds vs. 7-year-olds: $p = .219$; 6-year-olds vs. 7-year-olds: $p = 1.000$). For MAIN2, there was a significant difference between the 4-year-olds and all the age groups (4-year-olds vs. 5-year-olds: $p = .007$; 4-year-olds vs. 6-year-olds: $p = .002$; 4-year-olds vs. 7-year-olds: $p < .001$). The 5-year-olds differed significantly from the 7-year-olds but not from the 6-year-olds (5-year-olds vs. 6-year-olds: $p = 1.000$; 5-year-olds vs. 7-year-olds: $p = .014$). Finally, the oldest two age groups differed significantly from each other (6-year-olds vs. 7-year-olds: $p = .015$).

Next, the relationship between Swedish narrative comprehension and age (as a continuous variable) is analysed. Simple linear regression analyses showed that for both MAIN1 and MAIN2, age (in months) was a significant predictor (MAIN1: $F(1,95) = 20.92, p < .001, R^2 = .180$; MAIN2: $F(1,97) = 39.86, p < .001, R^2 = .291$). Similarly to the Arabic scores, the $R$-squared value for MAIN2 ($R^2 = .291$) is higher than that for MAIN1 ($R^2 = .180$) indicating a stronger relationship with age for MAIN2 narrative comprehension scores than that for MAIN1.

Age development (in form of a regression line) is visualized on a scatterplot for Swedish MAIN1 narrative comprehension scores (Figure 7.4) and for Swedish MAIN2 narrative comprehension scores (Figure 7.5).

For MAIN1, the pattern of the children’s scores in Swedish narrative comprehension resembles the Arabic results. About one third of the children ($N = 31$) scored below 50% (below 5 points) while about half of the children ($N = 52$) scored between 50%–90% (between 5 and 9 points). Seventeen children scored at ceiling with 10 points. The ceiling-scoring children belonged to all age groups (4-year-olds $N = 1$; 5-year-olds $N = 4$; 6-year-olds $N = 4$; 7-year-olds $N = 8$). For Swedish MAIN2, about half of the children ($N = 51$) scored below 50%; five of these children scored at floor level (zero points). The five floor-scoring children in Swedish MAIN2 comprehension were four 4-year-olds and one 5-year-old. Thirty-eight children scored between 50%–90%, while ten children scored at ceiling (5-year-olds $N = 3$; 6-year-olds $N = 1$; 7-year-olds $N = 6$). This pattern is also reminiscent of what was found for Arabic, although there were more extremes (floor and ceiling) in Swedish.
Figure 7.4. Scatterplot of Swedish MAIN1 (Cat/Dog) narrative comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

Figure 7.5. Scatterplot of Swedish MAIN2 (Baby Birds/Baby Goats) narrative comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

Next, children who scored the lowest in their age group in Swedish narrative comprehension will be discussed. In the 4-year-old group, BiAra4-15 scored at floor level (0 points) in both MAIN1 and MAIN2. (BiAra4-15 did not score high in either Arabic MAIN1 or MAIN2 with 3 points and 1 point, respectively, and remarkably low on both Arabic vocabulary measures.) As for
MAIN2, four 4-year-olds (including BiAra4-15) scored zero points. One of these children, BiAra4-02, scored 0 points on MAIN2 but managed to score 4 points on MAIN1. The child became highly distracted by the end of the testing session and was not fully concentrating when telling MAIN2 and answering the comprehension questions. The remaining two 4-year-olds who scored at floor level were BiAra4-03 and BiAra4-16. Both children had a late age of onset to Swedish (at age 4 and 3, respectively) and both scored below 50% (30 points) on both Swedish vocabulary (CLT) tasks. The reason why BiAra4-03 was able to perform the narration tasks (and manage to score 3 points on MAIN1) despite the late AoO to Swedish might be related to the child’s current language exposure which predominately (60%) occurs in Swedish.

In the 5-year-old group, BiAra5-20 scored low in both Swedish MAIN1 and MAIN2 with 2 points and 1 point, respectively. BiAra5-20 scored also low in Swedish vocabulary (CLT) production and was reported to have the majority of his language exposure in Arabic. BiAra5-20 had a relatively late AoO of Swedish (at age 3).

In the 6-year-old group, BiAra6-20 scored 2 points on both MAIN1 and MAIN2. This child was also mentioned in Section 6.3.2 since he scored remarkably low in both Swedish vocabulary (CLT) measures. BiAra6-20 was reported to have a late AoO of Swedish (at age 4). Furthermore, the child spoke only/mostly Arabic at home with parents and siblings and was estimated to be exposed to mostly (60%) Arabic throughout the day.

As for the 7-year-olds, BiAra7-17 scored low in MAIN1 (2 points) but a little higher in MAIN2 (4 points). BiAra7-17 did score at almost ceiling with 9 points in each of Arabic MAIN1 and MAIN2. Recall that BiAra7-17 scored low in both vocabulary production in Swedish (20/60) and Arabic (27/60), and was reported to have a late age of onset in Swedish (at around age 5) and daily language exposure to English in addition to Arabic and Swedish. Table 7.7 contains the answers (and their English translation) provided by BiAra7-17 (age 7;11) on Swedish MAIN1.

There were quite a few children in the 4-year-old group and the 5-year-old group that scored at ceiling (10 points) on MAIN1 and/or MAIN2. One ceiling-scoring child on MAIN1 was BiAra4-04 (the child also scored quite high on MAIN2 with 8 points). BiAra4-04’s Arabic narrative comprehension scores were not as high with 3 points on MAIN1 and 2 points on MAIN2. The child also scored high on Swedish vocabulary (CLT) comprehension (52/60) and production (41/60) but not as high on Arabic vocabulary comprehension (47/60) and production (22/60). BiAra4-04 was reported to be exposed the majority (60%) of her time to Swedish. Additionally, the parents reported that one of them spoke mostly Arabic with the child, while the other parent spoke Arabic and some German. BiAra4-04 had a relatively early onset to Swedish (at age 1) and high preschool attendance during the week.

---

188 On Swedish MAIN1, BiAra4-03 scored 3 points and BiAra4-16 scored 5 points.
Table 7.7. Answers from a low scoring 7-year-old (BiAra7-17, age 7;11) on Arabic MAIN1 (Cat) narrative comprehension questions.

<table>
<thead>
<tr>
<th>Question</th>
<th>Child’s answers</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.</td>
<td>till hämta den <em>(child points to butterfly)</em> ‘to fetch this’</td>
<td>1</td>
</tr>
<tr>
<td>D2.</td>
<td>&amp;eh (...) lamlade [: ramlade] i grassen <em>(child points to bush)</em>, jag vet inte</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>’&amp;eh (...) fell in the grass, I don’t know’</td>
<td></td>
</tr>
<tr>
<td>D3.</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>D4.</td>
<td>till hämta fisk</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>’to fetch fish’</td>
<td></td>
</tr>
<tr>
<td>D5.</td>
<td>håller en boll</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>’holds a ball’</td>
<td></td>
</tr>
<tr>
<td>D6.</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>D7.</td>
<td>till åta</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>’to eat’</td>
<td></td>
</tr>
<tr>
<td>D8.</td>
<td>jag vet inte, jag tror det att han ser katten åta</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>’I don’t know, I think that he sees the cat eat’</td>
<td></td>
</tr>
<tr>
<td>D9.</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>D10.</td>
<td>jag vet inte, nej, jag vet inte, jag säger ja, till hen hantat [: hämtat] fisk</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>till åta eller hämta en, en liten katt till komma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>’I don’t know, no, I don’t know, I say yes, to he brought fish to eat or bring a little cat to come’</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

*Note. Max = 10 points.*

Four children (three of them 5-years-old, BiAra5-02, BiAra5-09, BiAra5-11, BiAra7-03) scored at ceiling on both Swedish MAIN1 and MAIN2. All of these children also scored high on Swedish vocabulary (CLT) and were exposed to Swedish the majority (60%) of the time. (Only BiAra5-02 was reported to have an even amount of exposure to both languages during the day.) One of these ceiling-scoring younger children is BiAra5-09; the child also scored high on Swedish vocabulary (CLT) comprehension (57/60) and production (43/60).\textsuperscript{189} Similarly to BiAra4-04, BiAra5-09 had an age of onset to Swedish from age 1. Table 7.8 contains the answers that BiAra5-09 provided for the narrative comprehension of MAIN2.

\textsuperscript{189} On Arabic vocabulary (CLT), BiAra5-09 scored 47/60 on comprehension and 38/60 on production. For Arabic narrative comprehension of MAIN, the child scored 5/10 on MAIN1 and 4/10 on MAIN2. In other words, the child scored higher on Swedish than on Arabic, however, the child's Arabic scores were not at floor level. The child's parents reported that they spoke mostly Arabic with the child at home.
Table 7.8. Answers from a high-scoring 5-year-old (BiAra5-09, age 5;3) on Arabic MAIN2 (Baby Goats) narrative comprehension questions.

<table>
<thead>
<tr>
<th>Question</th>
<th>Child’s answers</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.</td>
<td>det är för att hjälpa barnet</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘it is because to help the child’</td>
<td></td>
</tr>
<tr>
<td>D2.</td>
<td>inte glad</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘not happy’</td>
<td></td>
</tr>
<tr>
<td>D3.</td>
<td>det är för att den har ramlat in i vattnet</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘it is because it has fallen into the water’</td>
<td></td>
</tr>
<tr>
<td>D4.</td>
<td>för att äta upp den där lilla fåren</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘in order to eat up that little sheep’</td>
<td></td>
</tr>
<tr>
<td>D5.</td>
<td>&amp;m ont</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>’&amp;m pain’</td>
<td></td>
</tr>
<tr>
<td>D6.</td>
<td>därför att fågel bit räven</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘because bird bit the fox’</td>
<td></td>
</tr>
<tr>
<td>D7.</td>
<td>därför att, därför att vargen vill äta upp &amp;eh getningen</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘because, because the wolf will eat up &amp;eh the goat/wasp’</td>
<td></td>
</tr>
<tr>
<td>D8.</td>
<td>glad</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘happy’</td>
<td></td>
</tr>
<tr>
<td>D9.</td>
<td>därför att han har räddat dom för att bita den här vargen (child points to fox in pic 6)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘because he has saved them because this wolf wanted to bite’</td>
<td></td>
</tr>
<tr>
<td>D10.</td>
<td>fågel, därför att fågel hjälpte dom (child points to pic 5)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘bird, because bird helped them’</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Max = 10 points.

To conclude, several low-scoring children on Swedish MAIN narrative comprehension also had low vocabulary scores on the Swedish CLTs, a late age of onset to Swedish and/or relatively low daily language exposure to Swedish. On the other hand, several of the high-scoring children also scored high on Swedish vocabulary and were exposed to Swedish during the majority of their day.

7.1.3.4 Comparison of Arabic and Swedish narrative comprehension development with age

For both languages, there is an increase in narrative comprehension scores with age. At group level, younger children score lower in narrative comprehension than older children in both Arabic and Swedish. This increase is more prominent for MAIN2 than for MAIN1. In Arabic, for both MAIN1 and MAIN2, the major increase with age (in the age-group mean score) happens between the 5-year-olds and the 6-year-olds, whereas in Swedish, the major increase happens between the 4-year-olds and the 5-year-olds.

As previously mentioned, both Arabic and Swedish MAIN1 and MAIN2 narrative comprehension scores significantly increased with age (in months).
Furthermore, comparing the \textit{R-squared} correlation values across the languages (Arabic \textit{MAIN1} $R^2 = .230$, \textit{MAIN2} $R^2 = .399$; Swedish: \textit{MAIN1} $R^2 = .180$, \textit{MAIN2} $R^2 = .291$), the Arabic \textit{R-squared} values are higher than the Swedish ones for both \textit{MAIN1} and \textit{MAIN2}. This indicates that Arabic narrative comprehension has a stronger relationship with age than Swedish narrative comprehension. However, on the individual level, children’s performances may be distributed differently for task and language.

Recall that for vocabulary (as measured by the CLT), both comprehension and production scores in both languages also significantly increased with age (in months). However, in contrast to the children’s narrative comprehension, vocabulary comprehension and vocabulary production scores in Swedish increased more strongly with age than what the Arabic vocabulary scores did.\textsuperscript{190}

\subsection*{7.1.3.5 Individual comprehension questions}

This section contains the children’s results on the individual comprehension questions for \textit{MAIN1} (Cat/Dog) and \textit{MAIN2} (Baby Birds/Baby Goats). In Table 7.9, the number (and percentage) of correct answers on the \textit{MAIN1} comprehension questions is provided. The \textit{MAIN} narrative comprehension questions for all narratives can be found in Appendix A1.4.

\begin{table}[h]
\centering
\caption{Number and percentage (%) of the correct answers on the \textit{MAIN1} (Cat/Dog) comprehension questions in Arabic and Swedish for all ages combined.}
\begin{tabular}{lll}
\hline
Question & Arabic (N = 99) & Swedish (N = 97) \\
\hline
D1. Episode 1 Goal & 88 (89\%) & 73 (75\%) \\
D2. Episode 1 IST & 60 (61\%) & 66 (68\%) \\
D3. Episode 1 \textit{IST rationale} & 49 (50\%) & 50 (52\%) \\
D4. Episode 2 Goal & 87 (88\%) & 72 (74\%) \\
D5. Episode 2 IST & 56 (56\%) & 69 (71\%) \\
D6. Episode 2 \textit{IST rationale} & 54 (54\%) & 64 (66\%) \\
D7. Episode 3 Goal & 95 (96\%) & 90 (93\%) \\
D8. Episode 3 \textit{IST Theory of Mind} & 51 (52\%) & 58 (60\%) \\
D9. Episode 3 \textit{IST rationale Theory of Mind} & 49 (49\%) & 56 (58\%) \\
D10. Overall plotline (Theory of Mind) & 43 (43\%) & 49 (51\%) \\
\hline
\end{tabular}
\textit{Note. IST = Internal State Term.}
\end{table}

As can be seen in Table 7.9, the pattern in the two languages was quite similar. In both languages, the goal questions (D1, D4, D7, for example D1: ‘Why does the cat/dog jump/leap forward?’) were the questions that were answered most correctly. Not surprisingly, the IST questions (D2, D5, D8, for example D2: ‘How does the cat/dog feel?’) gave slightly higher response accuracies than providing their rationale (D3, D6, D9, for example D3: ‘Why do you

\textsuperscript{190} While both Arabic and Swedish vocabulary comprehension and production scores increased with age, their \textit{R-squared} values in Swedish were higher in both comprehension and production than in Arabic (Recall that Arabic vocabulary: comprehension $R^2 = .248$, production $R^2 = .106$; Swedish vocabulary: comprehension $R^2 = .263$, production $R^2 = .213$).
think that the cat/dog is feeling [angry/disappointed/hurt]?’). The largest difference between the IST and its rationale was found for question D2 and D3. The least accurately answered question for MAIN1 was the overall plotline question (D10: ‘Will the boy be friends with the cat/dog? Why?’) where only 43% (Arabic) and 51% (Swedish) of the children answered correctly. The response accuracy patterns very much mirror what has been found in the literature on other populations.

Next, the children’s performance on the MAIN2 comprehension questions is analysed. Table 7.10 gives an overview of the correct answers (and percentage) for each of the ten comprehension questions in both languages for MAIN2 (Baby Birds/Baby Goats).

Table 7.10. Number and percentage (%) of the correct answers on the MAIN2 (Baby Birds/Baby Goats) comprehension questions in Arabic and Swedish for all ages combined.

<table>
<thead>
<tr>
<th>Question</th>
<th>Arabic (N = 99)</th>
<th>Swedish (N = 99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Episode 1 Goal/IST</td>
<td>54 (55%)</td>
<td>61 (62%)</td>
</tr>
<tr>
<td>D2. Episode 1 IST</td>
<td>46 (46%)</td>
<td>48 (48%)</td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
<td>24 (24%)</td>
<td>37 (37%)</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
<td>89 (90%)</td>
<td>82 (83%)</td>
</tr>
<tr>
<td>D5. Episode 2 IST</td>
<td>72 (72%)</td>
<td>61 (61%)</td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
<td>48 (49%)</td>
<td>42 (43%)</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
<td>64 (64%)</td>
<td>51 (51%)</td>
</tr>
<tr>
<td>D8. Episode 3 IST Theory of Mind</td>
<td>20 (20%)</td>
<td>34 (34%)</td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale Theory of Mind</td>
<td>14 (14%)</td>
<td>26 (26%)</td>
</tr>
<tr>
<td>D10. Overall plotline (Theory of Mind)</td>
<td>46 (46%)</td>
<td>46 (47%)</td>
</tr>
</tbody>
</table>

Note. IST = Internal State Term.

Just as for MAIN1, the children’s results in MAIN2 in both languages followed quite a similar pattern. For both languages (and similar to the pattern observed in MAIN1), the goal questions (D1, D4, D7; for example D1: ‘Why does the mother bird fly away?’/ ‘Why was the mother goat in the water?’) were the most correctly answered questions in each episode. For Episode 1 and Episode 3, the accuracy rates for goals were somewhat lower for MAIN2 than for MAIN1. The goal accuracy for Episode 2 was higher in MAIN2 than in MAIN1, and it was the question that most children answered correctly (D4: ‘Why is the cat climbing the tree?’/ ‘Why does the fox jump/leap forward?’) with 90% (Arabic) and 83% (Swedish) accuracy. The accuracy rate on the IST questions (D2, D5, D8, for example D2: ‘How do the [baby birds/baby goats] feel?’) was lower than that of the goal questions (lower by 8 to 44 percentage points), and the question querying the rationale provided for the ISTs (D3, D6, D9, for example D3: ‘Why do you think that the baby birds/baby goat are/is feeling [bad]?’) had lower accuracy rates than the IST questions (by 8 to 24 percentage points).

191 Recall that for question D10, the child is supposed to answer correctly on both parts of the question to be awarded 1 point.
points). Although question D5 had a lower accuracy rate than the preceding goal question in Episode 2, the accuracy rate was still high (72% for Arabic and 61% for Swedish). The question with the lowest accuracy rate was D9 (‘Why do you think that the dog/bird is feeling [bad]?’) at 14% (Arabic) and 26% (Swedish) accuracy. In the final question addressing the overall plotline (D10: ‘Who does the mother bird/goat like best, the cat/fox or the dog/bird? Why?’), a little less than half of the children (46% Arabic, 47% Swedish) answered correctly.

Next, the accuracy in answering goal questions is analysed. The comprehension of goals and the production of goals in the narratives is further analysed in Section 7.2.3.4.

In order to understand age development further of the comprehension of goals in MAIN1 and MAIN2, a combined accuracy value (in percent) was calculated for each age group. This means that, for every child, the three goal scores (D1, D4 and D7) were added to form one total value with a maximum of 3 points. For each age group, the mean score of all the children’s total goal score was calculated and then converted into percent. Figure 7.6 visualizes this proportion of correctly understood goals in the narrative comprehension questions. For the exact percentages, see Table A2.2 in Appendix A2.2.

![Figure 7.6](image)

**Figure 7.6.** Line chart of Arabic and Swedish goal comprehension mean scores (in percentage, %) for MAIN1 and MAIN2 for all age groups.

For MAIN1, when queried by a comprehension question, the proportions of goals expressed in Arabic were high throughout all age groups (from 89% accuracy at age 4 and 90% accuracy at age 7). For Swedish, comprehension of MAIN1 goals increased steeply from age 4 to age 5 (from 55% to 64%), but less strongly from age 5 to 7 (with an increase in accuracy from 79% at age 5 to 88% by age 7).
For MAIN2, Arabic goals increase sharply almost at a constant rate from age 4 (55% accuracy) reaching 82% accuracy at age 7. For Swedish, a sharp increase of goal accuracy occurred from age 4 (44%) to age 5 (66%), and from age 6 (66%) to age 7 (at 83%). Interestingly, at age 7, the children’s ability to express their understanding of narrative goals in both their languages and in both MAIN1 and MAIN2 seems to reach an equivalent high level. Goal accuracy level (combined for all 3 goals) at age 7 for both languages: Arabic MAIN1 90%, Arabic MAIN2 82%, Swedish MAIN1 88%, Swedish MAIN2 83%.

Hence, when observing the performance of each age group separately (rather than of the whole sample), different accuracy rates for the comprehension of goals between the age groups were found. Younger children and older children did not perform alike.

7.1.3.6 Narrative comprehension: Summary

At group level, the children’s performance in the two languages was quite similar concerning narrative comprehension of MAIN1 and MAIN2. Similarities were also found between the two languages regarding the children’s performance on individual comprehension questions. The comprehension of goals differed between the children’s age, language, and narrative task (MAIN1 and MAIN2). However, at group level, the comprehension of goals reached near identical accuracy levels on all narrative tasks in both languages by age 7. Response accuracy varied between different comprehension questions indicating that individual questions in MAIN1 and MAIN2 were not equally easy/difficult.

In general, the children performed significantly better in MAIN1 than in MAIN2 narrative comprehension. For both languages, more children scored at ceiling in MAIN1 than in MAIN2, and more children scored at floor (or near floor-level) in MAIN2 than in MAIN1. There was no significant difference between the children’s performance in the two different stories in MAIN1 (Cat and Dog) or the two different stories in MAIN2 (Baby Birds and Baby Goats). The children’s performance on the four MAIN narrative tasks (2 in each language) correlated with each other. All four narrative tasks increased significantly with age (in months). Furthermore, age (in months) was a significant predictor of all four MAIN comprehension tasks. A testing order effect for Swedish was found. Children who were tested their second session in Swedish scored better in Swedish than those whose first testing session was in Swedish. Finally, through observing individual cases of low-performing children, daily language exposure and vocabulary knowledge seem to have an effect on the children’s performance on narrative comprehension in the respective language.
7.2 Narrative production

In this section, the children’s narrative production is analysed. First, the specific research questions are stated (7.2.1), followed by a description of the data and the statistical analyses used (7.2.2). Then, the results (7.2.3) are presented in relation to age, language and story. Furthermore, the production of different story components and of story complexity is described. At the end of the section, the results concerning narrative production are summarized.

7.2.1 Research questions

The following specific research questions concerning narrative production are asked:

- How does narrative production develop with age in Arabic and Swedish?
- Do children perform similarly in their two languages with respect to narrative production?
- Do children perform differently in the two narrative production tasks (MAIN1 and MAIN2)?
- How do different macrostructure story components develop with age?
- How does story complexity develop with age?

7.2.2 Data and statistical analyses

7.2.2.1 Data

Out of the 100 children who participated in the cross-sectional study in Sweden, a few children were not able to perform one or several narration tasks. In Arabic, data from four narratives was not available, as two children (BiAra4-02 and BiAra5-06) did not tell MAIN1 and MAIN2. In the case of the four-year-old BiAra4-02, the child had great difficulty in concentrating during the session and answering in the target language Arabic. Consequently, the child did not do MAIN1 and his MAIN2 was terminated in mid-testing. Concerning the 5-year-old BiAra5-06, when consulting the video recordings, it was noticed that BiAra5-06 was a timid child who did not speak much during the whole Arabic session. The child merely pointed to the MAIN pictures without speaking when asked to narrate.

In Swedish, four narratives were missing as well. For MAIN1, one four-year-old, BiAra4-05, was shy and quiet during the whole narration task. BiAra4-05 did however speak shorter sentences while telling MAIN2, hence, the child’s MAIN2 was included. Similarly, another 4-year-old child, BiAra4-19, remained quiet during the testing of MAIN1. During MAIN2, BiAra4-19 felt more comfortable and told the story. Recall that the narration of MAIN1 (Cat/Dog) appears very early in the session, directly after the first task (non-word repetition task). Some children are timid at the beginning of the session, but overcome their shyness as the session proceeds. One 5-year-old, BiAra5-
13, did not concentrate well during the testing session and searched for distractions from performing the tasks (e.g. playing with stickers, looking at the camera). He remained quiet during most of the narration tasks of MAIN1 and MAIN2 with the exception of a few mumbles and pointing. Hence, the child was excluded from the analyses of Swedish narrative production.

Table 7.11 gives an overview of the number of stories told in each language in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats).

### Table 7.11. Number of stories by task (MAIN1 and MAIN2) and language.

<table>
<thead>
<tr>
<th></th>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAIN1</td>
<td>MAIN2</td>
</tr>
<tr>
<td>Potential narrations</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Not carried out/Excluded</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

#### 7.2.2.2 Statistical analyses

The p-value was held constant at $p < .05$ (two-tailed) for all statistical analyses. The same analyses were performed on both languages (Arabic and Swedish) separately. MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) were analysed separately.

To determine whether the children scored similarly on the different narratives of MAIN1 (Cat vs. Dog) and MAIN2 (Baby Birds vs. Baby Goats), independent samples t-tests were run. Paired samples t-tests were run to test whether there was a difference between the children’s performance in the two languages, to test whether there was a difference between MAIN1 and MAIN2 in both languages, and to test whether there was a difference between the two stories of MAIN1 (Cat and Dog) and the two stories of MAIN2 (Baby Birds and Baby Goats). Pearson correlations (two-tailed) were run to investigate whether the narrative production tasks correlated with each other.

For each language, age effects were explored. One-way ANOVAs were run to observe whether there was a significant difference between the age groups. Post hoc analyses (with Bonferroni correction) were used to identify which age groups differed significantly from each other. The relation between the narrative production scores and age (in months) was analysed with simple regression analyses. The regression analyses were visualized as regression lines on scatterplots showing the children’s individual scores in relation to age.

In addition to these statistical investigations of story structure total scores, macrostructural components and story complexity are reported descriptively in form of proportions.

For each macrostructural component, the production of actually produced components was calculated. For example, the component setting can only be produced once in a narration while goal can be produced three times in each narration (since each story has three episodes). This means that for setting, the maximum number of possible times is calculated by multiplying the number
of children in the respective age group by 1, whereas for goals, the maximum number of instances is calculated by multiplying the number of children in the respective age group by three. The formula for calculating the percentage for each age group is: (the number of actual components produced / maximum number of all possible components) x 100.

The same method was used to calculate the proportions of macrostructural story complexity types. The actual number of different sequences was divided by the possible maximum number of sequences (‘no sequence’, attempt-outcome (AO), goal-attempt/goal-outcome (GA/GO), and goal-attempt-outcome (GAO)) that were produced in each age group. Each story consists of three episodes; this means that the possibility of producing a sequence (or no sequence) is three per story for each child. The formula for calculating the percentage thus is: (the number of actual GAOs produced / maximum number of all possible GAOs) x 100. As the MAIN scoring protocol leaves it open how story complexity should be reported, this method of reporting was chosen following Lindgren (2018), Gagarina et al. (2019), Lindgren & Bohnacker (2022), and Bohnacker et al. (2022).

7.2.3 Results
7.2.3.1 Overview of the results
In this section, an overview of narrative macrostructure scores in Arabic and Swedish is given. Table 7.12 displays the mean, standard deviation (SD) and ranges for both languages in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) for all the children combined. The means were quite low and barely reached one third of the maximum score.

Table 7.12. Arabic and Swedish narrative production scores for MAIN1 and MAIN2, all ages combined (4;0–7;11).

<table>
<thead>
<tr>
<th></th>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.4</td>
<td>4.7</td>
</tr>
<tr>
<td>SD</td>
<td>2.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Range</td>
<td>0–10</td>
<td>0–12</td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.5</td>
<td>4.3</td>
</tr>
<tr>
<td>SD</td>
<td>2.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Range</td>
<td>0–11</td>
<td>0–11</td>
</tr>
</tbody>
</table>

Note. Max narrative production score = 17 points. BB = Baby Birds, BG = Baby Goats.

At group level, independent samples t-tests showed no significant difference in either of the two languages between the two stories in MAIN1 (Arabic: \(t(94.25) = -0.139, p = .890\); Swedish: \(t(94.48) = -0.839, p = .404\)) and the two stories in MAIN2 (Arabic: \(t(95.51) = 0.067, p = .947\); Swedish: \(t(89.07) = -0.112, p = .911\)). See Appendix A2.1 for a comparative table between the
scores of MAIN1 (Cat and Dog) and MAIN2 (Baby Birds and Baby Goats) for all ages combined and for both languages. Since there was no difference between the stories, all following results will be reported for MAIN1 and MAIN2, and not for all four stories separately.

Paired samples t-tests showed that there was no significant difference between the two languages’ MAIN1 scores ($t$(94) = 1.848, $p = .068$) but that there was a difference between the MAIN2 scores of the two languages where the Arabic MAIN2 scores were significantly higher than the Swedish MAIN2 scores ($t$(96) = 3.260, $p = .002$, $d = .331$, small effect size).

As for the difference between MAIN1 and MAIN2 scores in each language, paired samples t-tests showed that there was no difference between the children’s performance in neither Arabic ($t$(97) = -3.47, $p = .729$) nor Swedish ($t$(96) = 1.535, $p = .128$). This means that, unlike the narrative comprehension scores (recall Section 7.1.3.1), the children, as a group, did not score significantly higher on MAIN1 than on MAIN2.

Pearson correlations (two-tailed) showed that in both languages, MAIN1 and MAIN2 correlated positively with each other (Arabic: $r = .506$, $N = 98$, $p < .001$, strong correlation; Swedish: $r = .598$, $N = 97$, $p < .001$, strong correlation). As for correlations between the languages, Arabic MAIN1 and Swedish MAIN1 showed a positive but weak correlation ($r = .255$, $N = 95$, $p = .013$) while no correlation was found between Arabic MAIN1 and Swedish MAIN2 ($r = .130$, $N = 97$, $p = .206$). Arabic MAIN2 showed a positive but weak correlation with both Swedish MAIN1 ($r = .298$, $N = 95$, $p = .003$) and Swedish MAIN2 ($r = .229$, $N = 97$, $p = .024$).

Furthermore, Pearson correlations (two-tailed) showed that there were correlations between the narrative production and narrative comprehension scores in each of MAIN1 and MAIN2 in both languages (Arabic MAIN1: $r = .352$, $N = 98$, $p < .001$, Arabic MAIN2: $r = .566$, $N = 98$, $p < .001$; Swedish MAIN1: $r = .567$, $N = 97$, $p < .001$; Swedish MAIN2: $r = .734$, $N = 99$, $p < .001$). This result indicates that, in both languages, children who performed well in narrative production also performed well on narrative comprehension of the same task.

Independent samples t-tests revealed that there was no effect of the order of testing on the narrative production scores for neither Arabic (MAIN1: $t$(95.81) = 1.119, $p = .266$; MAIN2: $t$(91.85) = 1.193, $p = .236$) nor Swedish (MAIN1: $t$(94.26) = -1.605, $p = .112$; MAIN2: $t$(95.24) = -1.989, $p = .050$). This means that although there was a small difference in the mean scores between the children that were tested in a certain language first and those tested second, this difference was not significant. 192

192 When dividing the large sample into age groups, independent samples t-tests revealed that testing order was significant for the two younger age groups in Swedish MAIN2 where children who were tested in Swedish second scored higher than children tested in Swedish first (4-year-olds: $t$(19.94) = 2.541, $p = .019$, $d = 1.07$; 5-year-olds: $t$(17.09) = 3.301, $p = .004$, $d = 1.275$). In the 6-year-old group, however, a significant test order effect was found for Arabic MAIN2.
In Figure 7.7, the children’s narrative production scores for both languages across the age groups are visualized.

**Figure 7.7.** Bar graph over MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) production for Arabic and Swedish for all age groups. Error bars show ±1 SD.

In the 4-year-old, 5-year-old and 6-year-old groups, the group mean scores for Arabic MAIN1 and Arabic MAIN2 were higher than for the respective Swedish MAIN1 and Swedish MAIN2. Only in the 7-year-old group did the mean scores in Swedish exceed those in Arabic. The 4-, 5-, and 6-year-olds performed almost similarly in the two languages. One exception is the relatively high performance of the 6-year-olds in Arabic MAIN2. The 7-year-old group showed a very different pattern than all the other age groups. The 7-year-olds (as a group) performed virtually equally well on all tasks with slightly lower mean score on Arabic MAIN1.

In general, all the mean scores are generally low; all of them being below 50% of the total score (17 points). The highest mean is found in the 7-year-old group in Swedish MAIN2 (with 6.7 points) while the lowest mean is found in the 4-year-old group, also in Swedish MAIN2 (with 2.5 points).

Paired samples t-tests showed that in the 4-year-old group, the Arabic scores were significantly higher than the Swedish scores in both MAIN1 ($t(18)$ = 2.139, $p = .046$, $d = .491$, small to moderate effect size) and MAIN2 ($t(20)$ = 2.326, $p = .031$, $d = .508$, moderate effect size). The only other significant difference in performance between the two languages was in the 6-year-old

\[ t(26.88) = 2.994, p = .006, d = 1.107 \]

but none for the 7-year-olds in any language or narration task. The test order effect for the two younger age groups can be explained as a beneficial training effect in the children’s weaker language, Swedish. However, keep in mind that the sample sizes become relatively small when making these testing order divisions for each age group (between 9–15 children per age group).
group where the children scored significantly higher on Arabic MAIN2 ($t(28) = 3.261, p = .003, d = .606, \text{moderate effect size}$). In all other age groups and narration tasks, there were no significant differences between the two languages.

In the following two sections, Arabic narrative production (Section 7.2.3.2) and Swedish narrative production (Section 7.2.3.3) are analysed in detail in relation to age and story (MAIN1 and MAIN2). Examples of high- and low-performing children in each language are discussed further.

### 7.2.3.2 Arabic narrative production

Table 7.13 provides the mean, standard deviation (SD) and range for Arabic narrative production scores in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) for each age group.

**Table 7.13.** Arabic narrative production scores for MAIN1 and MAIN2 by age groups.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>MAIN1 (Cat/Dog)</th>
<th>MAIN2 (BB/BG)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>4.6 (3.3)</td>
<td>3.7 (2.3)</td>
</tr>
<tr>
<td>(N = 21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-year-olds</td>
<td>5.9 (2.3)</td>
<td>5.1 (2.6)</td>
</tr>
<tr>
<td>(N = 24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-year-olds</td>
<td>5.1 (2.6)</td>
<td>6.1 (2.5)</td>
</tr>
<tr>
<td>(N = 29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-year-olds</td>
<td>5.8 (2.0)</td>
<td>6.5 (2.9)</td>
</tr>
<tr>
<td>(N = 24)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Max narrative production score = 17 points. BB = Baby Birds, BG = Baby Goats.

As can be seen in Table 7.13, there is an increase in the scores for MAIN1 from age 4 to age 5 but no further increase in the mean score in the 6-year-olds and 7-year-olds. For the score variation (SD), there is a decrease in 1 point from the 4-year-olds to the 5-year-olds. The SD remains at around 2 points for the 6-year-old and the 7-year-old groups. For the score range, the highest score (10 points) is found in the 5-year-old, 6-year-old and 7-year-old groups. However, in the 4-year-old group, there were some nearly just as high-performing children scoring 9 points.

A different pattern is seen for MAIN2. There is a steady increase in mean scores from the 4-year-olds throughout to the 7-year-olds. Unlike in MAIN1, the SDs in MAIN2 increase slightly the older the age group but remain at around 2 points. The highest-scoring children belong to the oldest two age groups (the 6-year-olds and 7-year-olds) with 11 points. The highest-performing 4-year-old scored 8 points.

Comparing MAIN1 and MAIN2, the 4-year-olds and 5-year-olds had a higher mean score in MAIN1 than in MAIN2 while the 6-year-olds and 7-year-olds scored higher in MAIN2 than MAIN1. For the SD, the 4-year-olds
had a larger SD in MAIN1 than in MAIN2, while the 7-year-olds had a smaller
SD in MAIN1 than in MAIN2. For the 5-year-old and 6-year-old groups, the
SDs are almost identical.

The score ranges are almost identical for the 4-year-olds and 5-year-olds
for MAIN1 and MAIN2. For the 6-year-olds, the range in MAIN2 was slightly
narrower than the range in MAIN1. For the 7-year-olds, in MAIN1, the lowest
score was 3 while the lowest score in MAIN2 was 0.

In order to examine whether there was a significant difference between the
age groups, one-way ANOVAs were run. For MAIN1, there was no difference
between the age groups ($p = .264$) while for MAIN2, a significant difference
between the age groups was found ($F(3,94) = 5.46, p = .001, \eta^2 = .148$, large
effect size). Post hoc analyses (with Bonferroni correction) showed a signifi-
cant difference between the 4-year-olds and the two oldest age groups (4-year-
olds vs. 6-year-olds: $p = .007$; 4-year-olds vs. 7-year-olds: $p = .003$) but not
between the 4-year-olds and the 5-year-olds ($p = .384$). No other difference
between the remaining age groups was found (5-year-olds vs. 6-year-olds: $p
= .890$; 5-year-olds vs. 7-year-olds: $p = .419$; 6-year-olds vs. 7-year-olds: $p
= 1.000$).

Similarly, when the relationship between Arabic narrative production and
age was analysed as a continuous variable (in months), simple linear regres-
sion analyses showed that age was not a significant predictor of Arabic
MAIN1 ($F(1,96) = 2.14, p = .147, R^2 = .022$) but for Arabic MAIN2 ($F(1,96)
=16.32, p < .001, R^2 = .145$).

In the figures below, age development (or lack of it) is visualized in the
form of a regression line on a scatterplot for Arabic MAIN1 (Figure 7.8) and
Arabic MAIN2 narrative production (Figure 7.9).
Figure 7.8. Scatterplot of Arabic MAIN1 (Cat/Dog) narrative production scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

Figure 7.9. Scatterplot of Arabic MAIN2 (Baby Birds/Baby Goats) narrative production scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

For Arabic MAIN1, the majority of the children ($N = 85$) scored below 50% (below 8.5 points) on narrative production. The remaining children ($N = 13$) scored either 9 or 10 points and belonged to all age groups (4-year-olds $N = 3$; 5-year-olds $N = 4$; 6-year-olds $N = 3$; 7-year-olds $N = 3$). The range of scores was wide. Children scoring at floor level (0 points) can be found in all groups
except in the 7-year-old group (total $N = 5$, three 4-year-olds: BiAra4-04, BiAra4-05, BiAra4-16; one 5-year-old: BiAra5-25; one 6-year-old: BiAra6-26). Six children scored only 1 point, half of them were 4-year-olds (BiAra4-06, BiAra4-09, BiAra4-14) and the other half were 6-year-olds (BiAra6-08, BiAra6-10, BiAra6-11). Similarly for MAIN2, the majority of the children ($N = 84$) scored below 50%, while the remaining children ($N = 14$) scored between 9 and 11 points, however none of them belonged to the 4-year-old group (5-year-olds $N = 2$; 6-year-olds $N = 5$; 7-year-olds $N = 7$). The range of scores was also quite wide for all age groups. Three children scored at floor level, two 4-year-olds (BiAra4-14 and BiAra4-21) and one 7-year-old (BiAra7-12). The next to lowest scoring children (scoring 1 point) were one 5-year-old (BiAra5-16) and one 7-year-old (BiAra7-13).

Next, selected children who scored much lower or higher than their peers in their respective age group are discussed, starting with examples from the two oldest age groups. In the 6-year-old group, BiAra6-11 scored low on both Arabic MAIN1 (1 point) and MAIN2 (3 points), but higher on Swedish MAIN1 and MAIN2 (6 points on each task). The mean scores of the 5-year-olds and the 7-year-olds are near identical on Arabic MAIN1, however the mean score of the 6-year-olds is lower than the 5-year-olds’ (recall Table 7.13). In general, scoring no points or 1 point on MAIN1 should be quite uncommon for a 6-year-old. This was the case, however, in Arabic MAIN1 for BiAra6-26 (who scored 0 points) and BiAra6-08, BiAra6-10, BiAra6-11 (who scored 1 point each). These four low-performing children may be the reason why the mean score of the 6-year-olds in Arabic MAIN1 was lower than that of the 5-year-olds. Low-scoring BiAra6-26 was restless during the Arabic testing session but tried his best to tell the story (without managing to score any point). The child was also mentioned in Section 6.3.2.1 for scoring lower than his peers on Arabic vocabulary (CLT). The three remaining low-scoring 6-year-olds (BiAra6-08, BiAra6-10, BiAra6-11) code-switched to Swedish frequently when telling their stories. In the parental questionnaire, the three children were each reported to have one parent speak with them mostly in Arabic while the other parent spoke evenly in Arabic and Swedish. All the above mentioned 6-year-olds had the majority (60–80%) of their daily language exposure in Swedish.

BiAra7-12 is a 7-year-old child who scored the lowest amongst her same aged peers in both MAIN1 (3 points) and MAIN2 (0 points). The child had also scored particularly low in Arabic vocabulary production (18/60). BiAra7-12 was reported to have the vast majority (95%) of her daily language exposure in Swedish and her parents reported that they spoke 50% Arabic and 50% Swedish with the child. As for the child’s Swedish narrative production,

---

193 BiAra6-10 and BiAra6-11 have also been mentioned earlier in Section 6.3.2.1 for scoring low on Arabic vocabulary (CLT).
BiAra7-12 scored 6 points on both MAIN1 and on MAIN2. Table 7.14 shows the Arabic MAIN1 (Dog) narrative production of BiAra7-12.

**Table 7.14.** Arabic narrative production of a low-scoring 7-year-old (BiAra7-12; age 7;2) on MAIN1 (Dog).

<table>
<thead>
<tr>
<th>Component</th>
<th>Child’s narration</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>waḥad sabi ‘am bi ruḥ ma’ balon</td>
<td>0 time</td>
</tr>
<tr>
<td></td>
<td>‘one boy is going with a balloon’</td>
<td>0 location</td>
</tr>
<tr>
<td><strong>Ep 1:</strong> Attempt</td>
<td>elteñe elkalib ‘am bi fut bi elšajra</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>’the second dog is entering the tree’</td>
<td></td>
</tr>
<tr>
<td><strong>Ep 2:</strong> Attempt</td>
<td>elsabi ‘am bi šil elbalon</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>’the boy is removing the balloon’</td>
<td></td>
</tr>
<tr>
<td><strong>Ep 2:</strong> Outcome</td>
<td>ba’den šelo</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>’then he removed it’</td>
<td></td>
</tr>
<tr>
<td><strong>Ep 3:</strong> Outcome</td>
<td>elkalib ‘am yakul elakil</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>’the dog is eating the food’</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>


In the youngest two age groups, there were quite a few high-scoring children on Arabic narrative production. Below are examples of high-performing children from the 4-year-old and 5-year-old groups.

In the 5-year-old group, BiAra5-11 scored high on Arabic MAIN1 (10 points) but surprisingly quite low on MAIN2 (3 points). When consulting the video recording, it was noticed that the child was in a hurry to see the pictures of MAIN2 and to narrate the story. So the low score for MAIN2 could be due to the child skipping parts of the story while rushing through. BiAra5-11 also scored high on Swedish MAIN1 (9 points) and MAIN2 (8 points). The child was reported to speak ‘almost only Arabic’ with the parents who in their turn replied ‘mostly in Arabic’. BiAra5-11 had the majority of her daily language exposure (60%) in Swedish, and scored 33/60 on Arabic vocabulary production and slightly higher on Swedish vocabulary production (46/60).

There were several cases of children in the two youngest age groups (4-year-olds and 5-year-olds) who scored high on MAIN1 but much lower on MAIN2 (for example, BiAra4-18: 9 points on MAIN1 and 3 points on MAIN2, BiAra4-19: 9 points on MAIN1 and 4 points on MAIN2, BiAra5-03: 10 points on MAIN1 and 6 points on MAIN2). Upon closer inspection of the video recording for each child’s Arabic testing session, signs of tiredness can be seen as the children reach narration the MAIN2 task. For example, in the case of BiAra4-19, children in an adjacent room had started to play and speak loudly, which caused the child to speed up the narration asking whether she would soon join her peers. The above-mentioned young children scoring high on Arabic MAIN1 (BiAra5-11, BiAra4-18, BiAra4-19, BiAra5-03) had a relative late age of onset to Swedish (at around age 3). Two children (BiAra4-19

---

194 Recall that the children tell MAIN2 nearly at the end of the session. See Section 4.3.2.
and BiAra5-03) were exposed the majority (60%) of their time to Arabic while BiAra4-18 had an even exposure to both languages and BiAra5-11 was mostly exposed (60%) to Swedish throughout the day.

A high-performing 4-year-old (BiAra4-13) scored the highest in her age group in both MAIN1 (9 points) and MAIN2 (8 points). The child was reported to hear almost only Arabic from her parents and have the majority (80%) of her daily language exposure in Arabic. Table 7.15 contains the narrative macrostructure components produced by BiAra4-13 on the Dog story.

**Table 7.15.** Arabic narrative production of a high-scoring 4-year-old (BiAra4-13; age 4;0) on MAIN1 (Dog).

<table>
<thead>
<tr>
<th>Component</th>
<th>Child’s narration</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>awwal ši ken fi kalib, fara, kan fi fara kbire, kanit badda tfut bi alb ʾesāgara</td>
<td>0 time 1 location</td>
</tr>
<tr>
<td>Ep 1: Goal</td>
<td>elkalib baddu yakila</td>
<td>1</td>
</tr>
<tr>
<td>Ep 1: Attempt</td>
<td>kalib kan waraha, baʾden fatit ʾuqrī</td>
<td>1*</td>
</tr>
<tr>
<td>Ep 1: Outcome</td>
<td>baʾden elkalib fawwat rasu, em taraʾ rasu liʾannu rasu kbir</td>
<td>1</td>
</tr>
<tr>
<td>Ep 2: Attempt</td>
<td>baʾden ʾili elbalon bi elʾesāgara […] battal yiʾdir ytulu, w baʾden toliʾ bi elʾesāgara</td>
<td>1</td>
</tr>
<tr>
<td>Ep 2: Outcome</td>
<td>w talo</td>
<td>1</td>
</tr>
<tr>
<td>Ep 2: IST as R</td>
<td>w baʾden imbasat liʾannu ʾgebo</td>
<td>1</td>
</tr>
<tr>
<td>Ep 3: Goat</td>
<td>elkalib kan baddo yakul elkorvat [@s]</td>
<td>1</td>
</tr>
<tr>
<td>Ep 3: Outcome</td>
<td>elkalib aʾad yakul elkorvat [@s]</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Eq. = Episode. IST as R = Internal State Term as Reaction. Max = 17 points. * In scoring discussions with Ute Bohnacker, a consensus was reached that this expression ‘the dog was behind her’ would be considered clear enough as correct for Attempt since it was produced in context with a previously mentioned explicit goal ‘wanted to eat her’ and the mouse entering in a hurry/quickly.

BiAra4-13 scored very low in Swedish MAIN1 (2 points) and MAIN2 (0 points). Furthermore, BiAra4-13 scored double as high on her Arabic vocabulary (CLT) production than her Swedish vocabulary (CLT) production. On Arabic vocabulary (CLT) production, BiAra4-13 scored 37/60 points, and on Swedish, she scored 18/60 points.
### 7.2.3.3 Swedish narrative production

Table 7.16 provides an overview of the Swedish narrative production scores for both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) distributed per age group.

**Table 7.16.** Swedish narrative production scores for MAIN1 and MAIN2 by age groups.

<table>
<thead>
<tr>
<th></th>
<th>4-year-olds (N = 20/22)*</th>
<th>5-year-olds (N = 24)</th>
<th>6-year-olds (N = 29)</th>
<th>7-year-olds (N = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MAIN1</strong> (Cat/Dog)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.8</td>
<td>4.5</td>
<td>4.8</td>
<td>6.5</td>
</tr>
<tr>
<td>SD</td>
<td>1.5</td>
<td>2.6</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Range</td>
<td>1–6</td>
<td>0–9</td>
<td>1–8</td>
<td>2–12</td>
</tr>
<tr>
<td><strong>MAIN2</strong> (BB/BG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.4</td>
<td>3.7</td>
<td>4.2</td>
<td>6.7</td>
</tr>
<tr>
<td>SD</td>
<td>1.7</td>
<td>3.1</td>
<td>2.2</td>
<td>3.0</td>
</tr>
<tr>
<td>Range</td>
<td>0–6</td>
<td>0–11</td>
<td>0–8</td>
<td>0–11</td>
</tr>
</tbody>
</table>

*Note.* Max narrative production score = 17 points. BB = Baby Birds, BG = Baby Goats.

*Twenty 4-year-old children told MAIN1 and twenty-two 4-year-olds told MAIN2.

For both MAIN1 and MAIN2, the score variation (SD) for the 4-year-olds was the smallest and the range was the narrowest amongst the age groups. In MAIN1, the highest performing children were 7-year-olds (scoring 12 points) while the lowest scoring child was a 5-year-old (0 points). In MAIN2, there were children scoring at floor level in all age groups. The highest scoring children were in the 5-year-old group (1 child) and in the 7-year-old group (2 children) with 11 points each. Throughout the age groups, there is an increase in the mean scores for both MAIN1 and MAIN2.

One-way ANOVAS confirm a difference between the age groups for both MAIN1 ($F(3,93) = 11.77, p < .001, \eta^2 = .275$, large effect size) and for MAIN2 ($F(3,95) = 11.12, p < .001, \eta^2 = .260$, large effect size). The eta-squared ($\eta^2$) for MAIN1 and MAIN2 are roughly equal, however MAIN1 had a higher value suggesting that the age development is stronger for MAIN1 than MAIN2.

For MAIN1, post hoc analyses (with Bonferroni correction) showed that there was a significant difference between the scores of all age groups, except between the 5-year-olds and the 6-year-olds with $p = 1.000$ (4-year-olds vs. 5-year-olds: $p = .039$; 4-year-olds vs. 6-year-olds: $p = .009$; 4-year-olds vs. 7-year-olds: $p < .001$; 5-year-olds vs. 7-year-olds: $p = .009$; 6-year-olds vs. 7-year-olds: $p = .018$). For MAIN2, there was a significant difference between the 7-year-olds and the remaining age groups (4-year-olds vs. 7-year-olds: $p < .001$; 5-year-olds vs. 7-year-olds: $p = .001$; 6-year-olds vs. 7-year-olds: $p = .006$) but not between the other age groups (4-year-olds vs. 5-year-olds: $p = .545$; 4-year-olds vs. 6-year-olds: $p = .082$; 5-year-olds vs. 6-year-olds: $p = 1.000$).
When considering age as a continuous variable (in months), simple linear regression analyses showed that for both MAIN1 and MAIN2, age (in months) was a significant predictor (MAIN1: $F(1,95) = 31.81$, $p < .001$, $R^2 = .251$; MAIN2: $F(1,97) = 29.04$, $p < .001$, $R^2 = .230$). A comparison between the $R$-squared values of MAIN1 ($R^2 = .251$) and MAIN2 ($R^2 = .230$) suggests a somewhat stronger relationship with age (in months) for MAIN1 than MAIN2.

Figure 7.10 and Figure 7.11 illustrate the age development (in form of a regression line) on a scatterplot for Swedish MAIN1 narrative production and Swedish MAIN2 narrative production, respectively.

![Figure 7.10. Scatterplot of Swedish MAIN1 (Cat/Dog) narrative production scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

For Swedish MAIN1, the vast majority of the children ($N = 92$) scored below 50% (below 8.5 points) and only five children (one 5-year-old and four 7-year-olds) scored above 50%. The highest scoring child was a 7-year-old (BiAra7-07) with 12 points. Only one child (5-year-old BiAra5-14) scored at floor level (0 points). No 7-year-old child scored below 2 points.

A similar pattern can be seen for Swedish MAIN2, where the vast majority of the children ($N = 92$) scored below 50% and only seven children (one 5-year-old and six 7-year-olds) scored above 50% (the children scored between 9 and 11 points). The three highest scoring children were BiAra5-25, BiAra7-14 and BiAra7-22. The lowest scoring children (floor level, scoring 0 points) belonged to all age groups and were comprised of nine children (four 4-year-olds: BiAra4-14, BiAra4-15, BiAra4-16, BiAra21; two 5-year-olds: BiAra5-
18, BiAra5-20; two 6-year-olds: BiAra6-22, BiAra6-26; one 7-year-old: BiAra7-05). The 7-year-old group has the widest range (0–11 points).

![Figure 7.11. Scatterplot of Swedish MAIN2 (Baby Birds/Baby Goats) narrative production scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.

In what follows, examples are provided of low-scoring and high-scoring children in Swedish narrative production, starting with the two older age groups. In the 6-year-old group, BiAra6-20 and BiAra6-29 scored similarly on MAIN1 and MAIN2 with 3 points and 1 point, respectively. Both children have already been mentioned in Section 6.3.2.2 as low performers on the Swedish vocabulary (CLT) production task, and both children were reported to have had a relatively late age of onset (at age 4) to Swedish.

In the oldest age group, 7-year-old BiAra7-17 scored the lowest amongst his peers, with 2 points on MAIN1 and 1 point on MAIN2. This child has also been mentioned previously (in Sections 6.3.2.1 and 6.3.2.2) since he also scored relatively low in both Arabic and Swedish vocabulary (CLT) production. The child had a late age of onset to Swedish (age 5) and was reported to have had additional exposure to English at home, in addition to Arabic and Swedish. Furthermore, the child was not reported to being read to stories in Arabic at home but instead told stories once or twice a week. Table 7.17 contains the narrative macrostructure components produced by BiAra7-17 in the Baby Birds story.
Table 7.17. Swedish narrative production of a low scoring 7-year-old (BiAra7-17; age 7;11) on MAIN2 (Baby Birds).

<table>
<thead>
<tr>
<th>Component</th>
<th>Child’s narration</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>en fågel</td>
<td>0</td>
</tr>
<tr>
<td>Ep 1: Attempt</td>
<td>fågel flyga</td>
<td>0</td>
</tr>
<tr>
<td>Ep 1: Outcome</td>
<td>och en och fågel kommit med till äta</td>
<td>0</td>
</tr>
<tr>
<td>Ep 2: Attempt</td>
<td>sen katten går i träden [: trädet]</td>
<td>0</td>
</tr>
<tr>
<td>Ep 2: Outcome</td>
<td>sen katten litet så åtar [: åter] på en fågel</td>
<td>0</td>
</tr>
<tr>
<td>Ep 3: Attempt</td>
<td>och sen doggen [@e] går till &amp;eh lossade katten</td>
<td>0</td>
</tr>
<tr>
<td>Ep 3: Outcome</td>
<td>och sen doggen [@e] kommit och stoppat</td>
<td>1*</td>
</tr>
</tbody>
</table>

Total 1

Note. Ep. = Episode. Max = 17 points. * Lenient scoring, somewhat understandable from the utterance that the child means to say that the dog comes and stops (the cat from eating the birds).

Moving on to some high-performing children in the younger two age groups. Recall from Table 7.16 that the highest score achieved in the 4-year-old group was 6 points on both MAIN1 and MAIN2. BiAra4-25 scored 6 points on MAIN2 and 3 points on MAIN1. The child was reported to have an even exposure to both languages throughout her day and started to be exposed to Swedish from age 2. BiAra4-25 also scored 6 points on her Arabic narrative production of MAIN1 and 5 points on MAIN2. BiAra5-11 who was mentioned in Section 7.2.3.2 for having scored high on Arabic narrative production in MAIN1, scored also high in Swedish MAIN1 and MAIN2. BiAra5-25, another 5-year-old child who scored high on Swedish narrative production (7 points on MAIN1 and 11 points on MAIN2) did not score well on his Arabic (0 points on MAIN1 and 2 points on MAIN2). BiAra5-25 was reported to have his age of onset of Swedish at birth and have the majority (60%) of his daily language exposure in Swedish. Furthermore, the child was reported to have daily joint book reading sessions with his parents in Swedish (and not in Arabic). Table 7.18 provides an example of a high-scoring 5-year-old child in the narrative production of MAIN2 in Swedish.
Table 7.18. Swedish narrative production of a high-scoring 5-year-old (Bi Ara5-25; age 5;6) on MAIN2 (Baby Goats).

<table>
<thead>
<tr>
<th>Component</th>
<th>Child’s narration</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>det var en get som drunknade i vatten</td>
<td>0 time</td>
</tr>
<tr>
<td>Ep 1: Attempt</td>
<td>‘there was a goat that drowned in water’</td>
<td>1 location</td>
</tr>
<tr>
<td>Ep 1: Outcome</td>
<td>och räddade den</td>
<td>1</td>
</tr>
<tr>
<td>Ep 2: IST as IE</td>
<td>sen kom räven &lt;man ser att&gt; [?] han var sugen att åta upp dom</td>
<td>1</td>
</tr>
<tr>
<td>Ep 2: Goal</td>
<td>‘then the fox comes &lt;one sees that&gt; [?] he was keen to eat them up’</td>
<td></td>
</tr>
<tr>
<td>Ep 2: Attempt</td>
<td>sen kom räven &lt;man ser att&gt; [?] han var sugen att åta upp dom</td>
<td>1</td>
</tr>
<tr>
<td>Ep 2: IST as R</td>
<td>‘then the fox comes &lt;one sees that&gt; [?] he was eager to eat them up’</td>
<td></td>
</tr>
<tr>
<td>Ep 3: IST as IE</td>
<td>fägel där uppe och tittade</td>
<td>1</td>
</tr>
<tr>
<td>Ep 3: Attempt</td>
<td>och sen kom fågeln och bet räven</td>
<td>1</td>
</tr>
<tr>
<td>Ep 3: Outcome</td>
<td>‘and then the bird came and bit the fox’</td>
<td>1</td>
</tr>
<tr>
<td>Ep 3: IST as R</td>
<td>så hon var väl glad</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>

Note. Ep. = Episode, IST as IE = Internal State Term as Initiating Event, IST as R = Internal State Term as Reaction. Max = 17 points.

7.2.3.4 Individual macrostructure components

This section looks at the distribution and proportion of different macrostructure components in terms of language and narration task (MAIN1 and MAIN2). The macrostructural components are setting, internal state term as initiating event, goal, attempt, outcome and internal state term as reaction. Recall that in each story there are three episodes, which gives the child one opportunity to express setting (time and place) in the introduction and three opportunities to express each of the remaining components.

Starting with Arabic, Figure 7.12 (Arabic MAIN1, Cat/Dog) and Figure 7.13 (Arabic MAIN2, Baby Birds/Baby Goats) display the macrostructure components in form of bar charts with percentages (%) for each component type. For the exact percentages, refer to Table A2.4 in Appendix A2.3.1.

In MAIN1 (Figure 7.12), for all age groups, outcomes were the most produced component followed by attempts. For MAIN2 (Figure 7.13), it was the
other way around with attempts being the most produced followed by outcomes. The setting was expressed to a lower degree in MAIN1 in comparison to MAIN2, but for both narration tasks, the 7-year-olds scored much higher than the remaining age groups. The least frequent expressed component in both MAIN1 and MAIN2 was internal state terms as reaction.

In general, goals were expressed to a lesser degree than attempts and outcomes in both MAIN1 and MAIN2. These results do not mirror the children’s understanding of goals (recall Section 7.1.3.5) where the accuracy rates of the comprehension of goals were very high. Surprisingly, the production of goals seem to increase with age in MAIN2 but decrease with age in MAIN1. This is partly in line with the comprehension of goals where accuracy rates for MAIN2 also increased with age; however, the accuracy rate for goal comprehension remained very high in all age groups.

Effects of development in relation to age were more visible for the macrostructural components of MAIN2 than in MAIN1. In MAIN1, age improvements occurred in the setting and outcomes whereas in MAIN2, age improvement was visible in goals, attempts, outcomes and internal state terms as reaction. Surprisingly, in neither MAIN1 nor MAIN2 are the 7-year-olds always scoring higher than the remaining age groups. For example for MAIN1 goals, the 7-year-olds are scoring the lowest number of goals.

Figure 7.12. Percentage (%) of produced macrostructure component types in Arabic MAIN1 (Cat/Dog) for each age group. IST = Internal state term, IE = Initiating event, R = Reaction.
Figure 7.13. Percentage (%) of produced macrostructure component types in Arabic MAIN2 (Baby Birds/Baby Goats) for each age group. IST = Internal state term, IE = Initiating event, R = Reaction.

For Swedish, macrostructure component types (in percentages, %) are reported in bar charts in Figure 7.14 for MAIN1 (Cat/Dog) and Figure 7.15 for MAIN2 (Baby Birds/Baby Goats). For the exact percentages, refer to Table A2.5 in Appendix A2.3.1.

Similarly to Arabic, in Swedish MAIN1 (Figure 7.14), outcomes were most commonly produced followed by attempts. For MAIN2 (Figure 7.15), the results varied according to age group with either attempts or outcomes being the most commonly produced macrostructural component. For example, for the 7-year-olds, both attempts and outcomes were quite commonly produced, however, the outcomes to a slightly higher degree. For the 6-year-olds, it was the other way around. For the 5-year-olds, attempts were far more common than outcomes.

Also, similarly to Arabic, the Swedish MAIN1 setting was expressed less frequently than the MAIN2 setting, with the 7-year-olds scoring substantially higher than the other age groups in both MAIN1 and MAIN2 setting. The least commonly produced macrostructure component in Swedish MAIN1 and MAIN2 was internal state terms as reaction, similar to what occurred in Arabic. In general, goals were the least commonly produced component in comparison to attempts and outcomes. The children’s ability to produce goals did not reflect their ability to understand goals in Swedish MAIN1 and MAIN2 (recall Section 7.1.3.5 where the comprehension of goals had high accuracy rates and increased throughout the age groups).

In Swedish MAIN1, the provision of internal state terms as initiating event and attempts improved with age, while for Swedish MAIN2 increases with age were mostly visible for attempts and outcomes. For goals in MAIN1, the 4-year-olds had a lower provision rate than the remaining age groups (whose accuracy rates were near identical). For MAIN2, effects of development in
relation to age were also visible in the production of goals. An exception is seen in the 5-year-old group whose provision rate was the lowest among the all age groups. For almost all components (except for goals and internal state terms as reaction in MAIN1), the 7-year-olds had higher provision rates than the other age groups.

![Figure 7.14. Percentage (%) of produced macrostructure component types of Swedish MAIN1 (Cat/Dog) for each age group. IST = Internal state term, IE = Initiating event, R = Reaction.](image)

![Figure 7.15. Percentage (%) of produced macrostructure component types of Swedish MAIN2 (Baby Birds/Baby Goats) for each age group. IST = Internal state term, IE = Initiating event, R = Reaction.](image)

### 7.2.3.5 Story complexity

In this section, macrostructural story complexity is reported for each language and narration task (MAIN1 and MAIN2). Macrostructural complexity was measured in each episode by the occurrence of different types of sequences:
attempt-outcome (AO), goal-attempt/goal-outcome (GA/GO) and goal-attempt-outcome (GAO) i.e. full sequence. The lack of occurrence of any sequence was labelled as ‘no sequence’ and included not mentioning any component at all or mentioning any of the components alone (e.g. only one attempt and no other component in that episode). In each of the three episodes, a child has the opportunity to express one of the following AO, GA/GO, GAO or ‘no sequence’.

In Figure 7.16 (Arabic MAIN1, Cat/Dog) and Figure 7.17 (Arabic MAIN2, Baby Birds/Baby Goats), story complexity for Arabic is reported in the form of percentages (%) for each complexity type. For the exact percentages, refer to Table A2.6 in Appendix A2.3.2.

For Arabic MAIN1, there seems to be no clear pattern regarding development with age except for goal-attempt/goal-outcome (GA/GO) that decreases as the children become older. In contrast, in Arabic MAIN2, there is a gradual decrease in ‘no sequence’ in the oldest two age groups in comparison to the 4-year-olds and 5-year-olds, and a gradual increase in the proportion of full sequences (GAO) from the 4-year-old group to the 7-year-old group. The production of GAO in MAIN2 increased from 4.8% in the 4-year-old group to 19.4% in the 7-year-olds.

In both MAIN1 and MAIN2, however, the most frequent complexity type was ‘no sequence’ while the most common type of sequence was attempt-outcome (AO), except for the 4-year-olds where the children produced more goal-attempt/goal-outcome sequences (GA/GO).

Figure 7.16. Percentage (%) of story complexity types in Arabic MAIN1 (Cat/Dog) for each age group. AO = attempt-outcome, GA/GO = goal-attempt/goal-outcome, GAO = goal-attempt-outcome.
Next, Swedish story complexity is described. Figure 7.18 and Figure 7.19 show the distribution of sequence types in percentages (%) for each age group in Swedish MAIN1 (Cat/Dog) and Swedish MAIN2 (Baby Birds/Baby Goats), respectively. For the exact percentages, refer to Table A2.7 in Appendix A2.3.2. Unlike in Arabic, a clearer patterns of development with age can be seen for both Swedish MAIN1 and Swedish MAIN2 story complexity. The production of ‘no sequence’ was the most frequent complexity type and its proportional use decreased with age, while the use of attempt-outcome (AO) increased. Similarly to in Arabic, the most commonly produced sequence was AO for all age groups in both MAIN1 and MAIN2 except for the 4-year-olds who produced more goal-attempt/goal-outcome (GA/GO). As for the production of a full sequence (goal-attempt-outcome, GAO), the 4-year-old group scored very low in both Swedish MAIN1 (1.7%) and Swedish MAIN2 (1.5%). This percentage increased with age reaching 11.1% on MAIN1 and 15.3% on MAIN2 by age 7.
Table 7.19 provides examples of the different story complexity types that were produced by the children narrating in Arabic or Swedish. Translation equivalents in English are provided underneath.
### Table 7.19. Examples of story complexity types in Arabic or Swedish.

<table>
<thead>
<tr>
<th>Story complexity</th>
<th>Child code</th>
<th>Story / Ep. (language)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAO</td>
<td>BiAra7-23</td>
<td>MAIN1 (Cat) / Ep.3 (Ara)</td>
<td>elbazzuni šefet samak, badda takul elsamak, ba’den eği köm tokul elsamaki, elbazzuni ki takila elsamaki. 'the cat saw fish, she wants to eat the fish, then she came [and] grabbed the fish, the cat eats the fish.'</td>
</tr>
<tr>
<td>GA</td>
<td>BiAra5-17</td>
<td>MAIN2 (BB) / Ep.2 (Swe)</td>
<td>katten (jättearg) han skulle äta en fågel, (...) och sen / [:] sen katt klättra upp på träden. 'the cat (very angry) he wanted to eat a bird, (...) and then [:] then cat climbed up on the tree.'</td>
</tr>
<tr>
<td>GO</td>
<td>BiAra5-07</td>
<td>MAIN1 (Dog) / Ep.1 (Ara)</td>
<td>elkalib biddu yakul elfar, ba’den elfar harab tahit elsağara. 'the dog wants to eat the mouse, then the mouse escaped under the tree.'</td>
</tr>
<tr>
<td>AO</td>
<td>BiAra4-20</td>
<td>MAIN1 (Dog) / Ep. 2 (Swe)</td>
<td>(...) han hoppa (..) [took] ballong. '(...) he jump (..) then took [took] balloon.'</td>
</tr>
<tr>
<td>No sequence</td>
<td>BiAra6-13</td>
<td>MAIN2 (BB) / Ep.1 (Swe)</td>
<td>först säger hon att jag ska gå iväg ett litet promenad, sen går hon, sen kom hon tillbaks. 'first she says that I should go away for a little walk, then she goes, then she came back.'</td>
</tr>
</tbody>
</table>

**Note.** GAO = goal-attempt-outcome (full episode), GA = goal-attempt, GO = goal-outcome, AO = attempt-outcome, Ep. = episode, Ara = Arabic, Swe = Swedish, BB = Baby Birds, BG = Baby Goats.

#### 7.2.3.6 Narrative production: Summary

In general, the mean scores for Arabic and Swedish narrative production were low for all ages. There was a clear increase in scores with respect to age for Swedish MAIN1 and Swedish MAIN2 as well as for Arabic MAIN2, but not for Arabic MAIN1. The increase in the Swedish production scores was more evident than the increase in Arabic MAIN2; however, all of the scores significantly increased with age (in months). As for Arabic MAIN1, there was no significant difference between the age groups (and no correlation with age as a continuous variable).

In general, the scores in all age groups were widely scattered. In both Arabic and Swedish MAIN1 and MAIN2, only few children scored above 50% (8.5 points). For Arabic MAIN1, these high-performing children belonged to all age groups, while for Arabic MAIN2, the high-performing children belonged mostly to the oldest two age groups (6-year-olds and 7-year-olds). For Swedish MAIN1 and MAIN2, the high-performing children belonged to the 5-year-olds and the 7-year-olds, but the majority of these children were from the oldest age group.

At group level, there was no difference in narrative production scores between the different types of stories in MAIN1 (Cat and Dog) and in MAIN2.
(Baby Birds and Baby Goats) in either language. Moreover, the two languages’ MAIN1 and MAIN2 scores correlated with each other, meaning that children who performed well on MAIN1 also performed well on MAIN2 in that language. Correlations across languages were either non-existent (between Arabic MAIN1 and Swedish MAIN2) or weak (between Arabic MAIN1 and Swedish MAIN1, Arabic MAIN2 and Swedish MAIN1, Arabic MAIN2 and Swedish MAIN2). This means that children who scored high on Arabic narrative production did not necessarily score high on Swedish narrative production. There was no significant difference between the children’s performance in either language for MAIN1, but there was for MAIN2 where the children (at group level) scored significantly higher in Arabic. However, for both languages, there was no significant difference between the narrative production scores of MAIN1 and MAIN2. Hence, there was no task effect for narrative production, unlike for narrative comprehension, where children scored significantly higher on MAIN1 than on MAIN2 in both languages.

Individual cases of high and low performing children showed a possible relation between AoO, current language exposure and vocabulary scores with the children’s performance on narrative production in the respective language.

As for the children’s performance on the production of different macrostructural components, the results were presented in percentage form for each component type (at age group level). For Arabic, in MAIN1, the most commonly produced component type was outcomes while for MAIN2 it was attempts. For both MAIN1 and MAIN2, the least produced component type was internal state terms as reaction. Setting and internal state terms as initiating event were more commonly produced in MAIN2 that in MAIN1. MAIN2 showed clear effects of development with age concerning goals, attempts, outcomes and internal state terms as reactions. For MAIN1, development with age was mostly visible in setting and outcomes. For Swedish, in MAIN1 the most commonly produced component type was outcomes (similar to Arabic MAIN2). In MAIN2, the most commonly produced component types were attempts and outcomes. Similarly to Arabic, the least produced component type was internal state terms as reaction. The children produced more commonly setting and internal state terms as initiating event in MAIN2 in comparison to in MAIN1. Interestingly, for both MAIN1 and MAIN2, the setting component produced by the oldest age group (7-year-olds) was more than double that of each of the remaining age groups. In Swedish MAIN1, a clear increase with age was found for the production of internal state terms as initiating event and attempts while for MAIN2, age effects were most prevalent for attempts and outcomes.

The results for macrostructural complexity were also presented in percentage form for each language and narration task (at age group level). For both languages, the production of ‘no sequence’ was the most frequent. For Arabic MAIN1 and MAIN2, the most commonly produced complexity type was attempt-outcome (AO) for all age groups except for the 4-year-olds where the
children produced more goal-attempt/goal-outcome (GA/GO). In Arabic, for MAIN1 there seemed to be no clear trend while for MAIN2 there was a gradual increase of goals with age. For Swedish, both MAIN1 and MAIN2 showed a similar pattern with a gradual increase of GAOs (full sequences) and a decrease of ‘no sequences’ with age. Similarly to Arabic, attempt-outcome (AO) was the most commonly produced complexity type for all age groups except for the 4-year-olds who most commonly produced goal-attempt/goal-outcome (GA/GO).

7.3 Discussion
The results of narrative comprehension (Section 7.3.1) and narrative production (Section 7.3.2) are discussed in relation to previous literature.

7.3.1 Narrative comprehension
In this chapter, the following research questions on narrative comprehension were asked:
- How does narrative comprehension develop with age in Arabic and Swedish?
- Do children perform similarly in their two languages with respect to narrative comprehension?
- Do children perform differently in the two narrative comprehension tasks (MAIN1 and MAIN2)?
- Do children perform differently on different types of narrative comprehension questions?

7.3.1.1 Age effects
An increase in the mean scores for narrative comprehension was visible for both languages on both tasks (MAIN1 and MAIN2). At group level, children’s narrative comprehension measurably improves with age. This is in line with previous studies on narrative comprehension using MAIN for other language combinations (e.g. Bohnacker, 2016; Maviş et al., 2016; Roch et al., 2016; Lindgren, 2018; Öztekin, 2019; Bohnacker et al., 2020; Lindgren & Bohnacker, 2020) and studies using other narration stimulus materials with monolingual children (e.g. Trabasso et al., 1992; Berman & Slobin, 1994; Trabasso & Rodkin, 1994; Mäkinen, 2014).

The children’s scores in both languages increased significantly with age, and the $R$-squared values were higher in both languages for MAIN2 than for MAIN1 indicating that age was a stronger predictor of MAIN2 than for MAIN1. This was also found by Lindgren (2018), Bohnacker et al. (2020), Lindgren and Bohnacker (2020) who studied the narrative comprehension of
Swedish-speaking bilingual children (with other L1s) in Sweden. In general, answering the comprehension questions of MAIN1 was easier than MAIN2 (as shown by the significant difference between the children’s MAIN1 and MAIN2 scores. Because the MAIN1 comprehension scores were higher irrespective of age, there was less ‘room’ for an increase in scores with age. In the present study, more younger children scored at ceiling or near ceiling in MAIN1 in both languages, whereas for MAIN2, high scores were almost only achieved by the older children. The younger children scored quite low on MAIN2 in both languages. These findings echo those of Bohnacker et al. (2020) and Lindgren and Bohnacker (2020) for Turkish-Swedish, English-Swedish and German-Swedish.

However, unlike Bohnacker et al. (2020), Lindgren (2018) and Lindgren and Bohnacker (2020), the Arabic-Swedish-speaking children in the present study (as a group) showed a clearer age development for narrative comprehension in L1 (Arabic) than in L2 (Swedish). More younger children scored high on Swedish at an early age (as early as 4), however, there were also many low-performing children (from all four age groups) on the Swedish tasks (recall Figure 7.4 and Figure 7.5 in Section 7.1.3.3). In other words, the Swedish narrative comprehension scores were somewhat more scattered than the Arabic scores (recall Figure 7.2 and Figure 7.3 in Section 7.1.3.2).

A possible explanation as to why age development in Swedish was not as clear as in Arabic could be related to the characteristics of the bilingual group in the present study. Recall that nearly all the children had an AoO of Arabic from birth, but that there was great variation as to when the children started to hear Swedish regularly. This was not the case in Bohnacker et al. (2020) where the majority (83%) of the Turkish-Swedish-speaking children had an AoO of Swedish before the age of 3 in comparison to 48% in the present study. Furthermore, in Bohnacker et al. (2020), the vast majority of the children (92%) were born in Sweden, in contrast to 55% of the children in the present study. Many of the Arabic-Swedish-speaking children (in all age groups) were relative newcomers to Sweden and might not have had the opportunity to start their education in Swedish medium (pre)schools directly upon arrival. In Bohnacker et al. (2020), the 6- and 7-year-olds’ group mean scores in both MAIN1 and MAIN2 were higher in both languages than their same aged peers in the present study. More children scored at ceiling among the older Turkish-Swedish-speaking children than the Arabic-Swedish-speaking children, where only few children scored at ceiling, especially in MAIN2. Bohnacker et al. (2020) explain the visible increase in L2 (Swedish) by the fact that the majority of the Turkish-Swedish-speaking children had attended Swedish-medium (pre)schools and that this had boosted their Swedish narrative abilities at age 6–7. As a result, the 6- and 7-year-olds scored more uniformly to each other. For L1 (Turkish), language input was believed to be much more affected by the parents’ initiatives at home, which might vary from one family to another. The fact that the Arabic-Swedish-speaking group was not as homogeneous as
the Turkish-Swedish-speaking group in terms of AoO of Swedish might have been a contributing factor behind the less clear effect of age on Swedish than on Arabic narrative comprehension.

In sum, in the present study, overall narrative comprehension scores increased significantly with age in both languages, though more so in MAIN2 than in MAIN1. This is in line with the previous literature on the effect of age on children’s performance in other language combinations, and is documented here for the first time for a (relatively) large Arabic-speaking bilingual group. Unlike findings from other studies using MAIN in Sweden with other child bilingual populations of the same age range, narrative comprehension scores for the Arabic-Swedish-speaking children increased more strongly with age in their L1 (Arabic) than in their L2 (Swedish).

7.3.1.2 Language effects
At group level, no language effects were visible in the present study. The children scored similarly in Arabic and Swedish narrative comprehension in MAIN1 and in MAIN2. This is in line with several studies that have used MAIN with other language combinations (e.g. Bohnacker, 2016; Bohnacker et al., 2020; Lindgren & Bohnacker, 2020; Fiani et al., 2020; Gagarina et al., 2020; Kunnari & Vällimaa, 2020). When the sample was analysed with respect to age groups, no language effects were found there either for MAIN1 and MAIN2.

Furthermore, the narrative comprehension scores in Arabic and Swedish correlated positively with each other, which is in line with other studies (e.g. Bohnacker, 2016; Lindgren, 2018; Öztekin, 2019; Fiani et al., 2020). This may not come as a surprise since comprehension tasks test inferential understanding tapping into general cognitive skills and are not as much dependent on a particular language. Narrative comprehension skills concerning macrostructure are believed to transfer between or manifest similarly in both languages of bilingual children (Bohnacker & Gagarina, 2020).

However, as seen in examples of the individual children (Section 7.1.3.2 and Section 7.1.3.3), many of the low-scoring children had several features in common: late AoO to Swedish, low vocabulary (CLT) scores, and/or relatively low daily language exposure to the respective language. The effect of background factors on narrative comprehension is not analysed statistically in the present thesis; however, when examining the background characteristics of low/high-performing children, certain background (AoO, vocabulary performance, language exposure) factors repeatedly appear.

Bohnacker et al. (2020) found that, for 100 Turkish-Swedish-speaking children, expressive vocabulary (CLT) was a significant predictor of narrative comprehension for both MAIN1 and MAIN2 in L1 (Turkish) and MAIN2 in L2 (Swedish). (No effect of receptive vocabulary was found on any narrative task except for Swedish MAIN1.) By contrast, Lindgren and Bohnacker
(2020) found that, for 46 German-Swedish-speaking children (4–6), L2 (Swedish) expressive vocabulary (CLT) was not a significant predictor of Swedish MAIN1 and MAIN2, L1 (German) expressive vocabulary was a significant predictor of German MAIN1 and MAIN2. The authors speculate that the effect of expressive vocabulary is largely dependent on the children’s overall level of language skill and age.

After all, the assessment of a child’s understanding of the MAIN stories is done via analysing the child’s verbal expressions. Not having sufficient vocabulary to express oneself might influence the assessment outcome. Additionally, if a child does not have a basic understanding of the language, then s/he will not be able to understand the comprehension question (and as a result not be able to answer it). Future analyses of the present data should study the relationship between the Arabic-Swedish-speaking children’s vocabulary and narrative comprehension.

As for the order of testing (i.e. which language was tested first), the children performed better on Swedish narrative comprehension in both MAIN1 and MAIN2 when Swedish was the language tested second. No such testing order effect was found for Arabic.

Bohnacker et al. (2020) only found a training effect for Swedish MAIN1 for the 100 Turkish-Swedish-speaking children (4–7). The authors suggest that this finding may be related to the testing situation in the Swedish session where the child encounters an unfamiliar experimenter for the first time and is asked to tell a story in a format that s/he is not used to, in the child’s weaker language (Swedish). By the time the child tells MAIN2 in Swedish, the child would have already become familiar with the narration task. No testing order effect was found for MAIN1 or MAIN2 in Turkish, which was most of the children’s stronger language. Lindgren and Bohnacker (2020) found no training effect on Swedish (L2) scores of 46 German-Swedish-speaking children (age 4–6). However, a training effect was found for German (L1) which was the weaker language for many of the participants. Children whose first testing was in Swedish scored higher on German than children whose first testing was in German. Lindgren and Bohnacker explain this by the children’s lower language proficiency in German than in Swedish (as estimated by the parents and measured on vocabulary CLT tasks). Children who were tested second in German already knew what to expect from the narration task and comprehension questions, and as a result, were able to perform better in German.

In the present study, more children were rated to have a better proficiency in Arabic than in Swedish (recall Table 5.4 in Section 5.1.2), especially the younger children who were exposed to Arabic more than Swedish (recall Table 5.18 in Section 5.1.8). Hence, having been tested first in Arabic, the stronger language according to parental report, might have helped children with lower proficiency in Swedish to become more familiar with the task, and consequently score higher in Swedish than children whose first testing session was in Swedish.
7.3.1.3 Task and story effects

The current study showed that there was a task effect for narrative comprehension in both languages, where children scored higher on MAIN1 (Cat/Dog) than on MAIN2 (Baby Birds/Baby Goats). Other studies using MAIN have found the same task effects for other languages and language combinations (e.g. Swedish, German-Swedish, Turkish-Swedish in Lindgren, 2018; Turkish-Swedish in Bohnacker et al., 2020; English-Swedish in Bohnacker & Lindgren, 2020; German-Swedish in Lindgren & Bohnacker, 2020; Finnish and Finnish-Swedish in Kunnari & Välimaa, 2020). Hence, the author agrees with Bohnacker et al. (2020), Bohnacker and Lindgren (2020), Lindgren (2018) and Lindgren and Bohnacker (2020) that when assessing a child’s narrative comprehension using MAIN, a child’s performance on Cat/Dog cannot be straightforwardly compared with her/his performance on Baby Birds/Baby Goats, even when both narration tasks were administered in the same mode.195

One explanation for why the children performed better on MAIN1 than on MAIN2 may be related to the time/order that the tasks are performed. Recall that MAIN1 is performed early in the testing session whereas MAIN2 is performed towards the end when the children might be tired or bored (Lindgren, 2018; Wehmeier, 2020). However, this would mean that children are expected to perform lower in MAIN2 (than in MAIN1) on narrative production as well (which was not the case, recall Section 7.2.3.1). Bohnacker et al. (2020) instead argue that children are able to answer the majority of the questions of MAIN2 less accurately than the questions of MAIN1 because the MAIN2 questions are in general more challenging than the MAIN1 questions.

As for story effects, the children in the present study scored equally well on both stories of MAIN1 (Cat and Dog) and on both stories of MAIN2 (Baby Birds and Baby Goats) in both languages. This was not the case in Bohnacker et al. (2020) where the Turkish-Swedish-speaking children performed better on the narrative comprehension of Baby Goats than Baby Birds.196 The authors speculate that this could be due to the different visual cues needed to answer questions D2 (internal state) and D3 (internal state rationale) in the two stories.197 The authors describe several pictorial differences between Baby Goats

---

195 Gagarina et al. (2012, 2015) had originally constructed the MAIN picture sequences and stories to be structurally parallel and assumed that they also were equally difficult, where Baby Birds/Baby Goats were intended to be used for telling mode and Cat/Dog were intended to be used for retelling mode. Studies that have used this narrative elicitation mode (e.g. Kunnari et al., 2016; Maviş et al., 2016; Roch et al., 2016) found that children performed better on Cat/Dog than in Baby Birds/Baby Goats but attributed the difference in performance to the elicitation mode (retelling vs. telling) rather than to task differences.

196 Similar differences in performance between Baby Birds and Baby Goats were also found by Lindgren (2018, 2019) and Bohnacker and Lindgren (2021). No differences were found between Cat and Dog in any of the studies, similarly to the present study.

197 Bohnacker et al. (2020) explain that, in Baby Goats, the facial expressions of the baby goat in the water are worried, likewise the face of its mother who is jumping towards it to rescue it.
and Baby Birds that may have contributed to the difference in response accuracies in D2 and D3 between the two stories.

As previously mentioned, no statistically significant difference was found in the present study between the children’s performance on Baby Birds and Baby Goats. However, closer inspection of the scores of the individual questions for Baby Birds and Baby Goats, a higher accuracy rate was indeed found for Baby Goats than Baby Birds on questions D2 and D3.  

A possible reason why no significant overall difference in comprehension between the two stories was found might be due to better performance in Baby Birds on other questions. For example, the children scored higher on Baby Birds than on Baby Goats on questions D1 (goal) and D10 (theory of mind) in Arabic, and question D7 (goal) in Swedish. Another reason behind the non-significant difference between Baby Birds and Baby Goats in the current study could be related to the language testing order. Recall that children who were tested second in Swedish scored higher on Swedish MAIN1 and MAIN2 in comparison to children who were tested in Swedish first. Hence, children who were tested second in Swedish might have told the Baby Birds story. Future research could make deeper quantitative and qualitative analyses of the children’s answers and further examine why the children scored better on Baby Birds on particular questions (and in particular languages).

At age 7, children scored close to ceiling (82%–90%) for the comprehension of goals, which accords well with previous reports in the literature on the early comprehension of goals (e.g. Bohnacker, 2016; Bohnacker & Lindgren, 2021; Fiani et al., 2020; Kunnari & Välimaa, 2020). In all episodes of MAIN1 and MAIN2, and in both languages, the comprehension of goals was less difficult than the comprehension of internal states or theory of mind. Similar results were found by Fiani et al. (2020) for (Lebanese) Arabic-French-speaking bilinguals (4–9). Bohnacker et al. (2020) and Lindgren and Bohnacker (2020) found that the comprehension of goals was very accurate in 4–7-year-old Turkish-Swedish-speaking and German-Swedish-speaking children, respectively, however, that the accuracy rates for internal states were at times higher than that of goals (especially in episode 2). Future studies could look deeper into why Arabic-speaking children in the present study were able to identify the goal of episode 2 better than other Swedish-speaking children. In general, however, the results of the present study are in line with previous studies that

---

198 For Arabic, accuracy rate in D2 Baby Birds = 33%, Baby Goats = 62% and accuracy rate in D3 Baby Birds = 16%, Baby Goats = 32%. For Swedish, accuracy rate in D2 Baby Birds = 33%, Baby Goats = 65% and accuracy rate in D3 Baby Birds = 25%, Baby Goats = 50%.

199 For Arabic, accuracy rate in D1 Baby Birds = 65%, Baby Goats = 43% and accuracy rate in D10 Baby Birds = 51%, Baby Goats = 40%. For Swedish, accuracy rate in D7 Baby Birds = 55%, Baby Goats = 47%.
show that the comprehension of goals increases with age and is at ceiling by age 6 or 7.

7.3.2 Narrative production
The following research questions were asked with regard to narrative production:
- How does narrative production develop with age in Arabic and Swedish?
- Do children perform similarly in their two languages with respect to narrative production?
- Do children perform differently in the two narrative production tasks (MAIN1 and MAIN2)?
- How do different macrostructure story components develop with age?
- How does story complexity develop with age?

7.3.2.1 Age effects
The effect of age was not the same for Arabic and Swedish narrative production scores. For Arabic, no age effects were found for MAIN1 (no significant increase with age and no significant difference between the age groups) while there was an increase for MAIN2 with age (where significant differences were only found between the youngest age group and the two oldest age groups). For Swedish, however, age was a significant predictor of both MAIN1 and MAIN2 production. Hence, there was a much clearer age development for Swedish than for Arabic.

The results in the present study are partly in line with other studies that have used MAIN to study narrative production. For example, Bohnacker et al. (2022) found a significant relationship between age (in months) and narrative production scores in both languages of Turkish-Swedish-speaking bilingual children (N = 100, 4–7) but also that the relationship with age was (slightly) stronger in the majority language (Swedish, L2) than in the home language (Turkish, L1). Similarly, Lindgren & Bohnacker (2022) found greater improvement with age in the majority language (Swedish, L2) compared to the home language (German, L1) of 46 German-Swedish-speaking children.

Bohnacker (2016) found a significant increase with age (5–7) in both languages of 52 English-Swedish-speaking bilingual children. Similarly, Gagarina (2016) found an increase in both languages of 58 Russian-German-speaking bilinguals between preschoolers (mean age 3;9) and first graders (mean age 7;0).

The fact that there was no increase with age for Arabic MAIN1 is thus not in line with previous studies. The Arabic-Swedish-speaking children’s narrative production scores on Arabic MAIN1 are scattered (recall Figure 7.8 in Section 7.2.3.2). Although the standard deviation decreased with age, the
score ranges were very wide in all age groups. Several 4-year-olds and 5-year-olds scored as high as the highest 6-year-olds and 7-year-olds, and several children from the 6-year-old group scored at floor level. This was not the case in Arabic MAIN2 where clearer age effects were visible.

As for the whole sample, the mean narrative production scores were low for all age groups. Most of the children scored below 50% on both narration tasks in both languages. The children who scored the highest were mostly 7-year-olds. In studies of Turkish-Swedish-speaking and German-Swedish-speaking children, Bohnacker et al. (2022) and Lindgren (2018) also found that the highest scores were produced by the oldest age groups. An exception to this pattern is the children’s performance on Arabic MAIN1 where several children from the younger age groups scored as high as the older children.

When examining whether testing order had any effect on MAIN1 or MAIN2 production, no significant effects were found for the whole sample, meaning that children who were tested in a language second did not score higher than children who were tested in that language first.

In sum, at group level, the children’s narrative production improved more strongly with age in the majority language (Swedish) than in the home language (Arabic). This finding echoes the results observed for the children’s vocabulary scores that increased with age in both languages but somewhat more strongly in Swedish than in Arabic. Observe, however, that the children’s narrative comprehension scores also improved with age but somewhat more prominently in Arabic than in Swedish.

7.3.2.2 Language effects

At group level, the children in the present study performed equally well in both languages in MAIN1, however, in MAIN2, the children scored significantly higher in Arabic than in Swedish. When investigating each age group alone, differences were only found in the 4-year-old group who performed significantly better in Arabic in both MAIN1 and MAIN2, and the 6-year-old group who scored significantly higher on Arabic MAIN2.

Recall that many of the 4-year-olds were reported (in the parental questionnaire) to have Arabic as their stronger language and that many of the younger children were exposed to Arabic for the majority of the day. Furthermore, while all of the 4-year-olds had an AoO of Arabic at birth, AoO of Swedish varied greatly among the children. Hence, it comes as little surprise that the youngest age group scored significantly better in the home language, their strongest language, on both narrative production tasks.

The only other difference observed between the languages was in MAIN2 for the 6-year-olds who scored significantly higher in Arabic. As a group, the 6-year-olds had a mean AoO of Swedish at age 2.7 where several children (N = 9, i.e. 31% of the 6-year-olds) were reported to having had only 2 years or
less of regular exposure to Swedish before the testing session. One explanation as to why the children scored so well on Arabic MAIN2 could be due to testing order and a training effect. Telling Arabic MAIN1 and answering its comprehension questions first might have given the children a hint of what to expect on Arabic MAIN2. Although there was no training effect for the sample as a whole on MAIN1 and MAIN2, when examining effects of testing order for each age group separately, an order effect was found for the 6-year-olds in Arabic where children who were tested second in Arabic scored higher on Arabic MAIN2 (but not higher on Arabic MAIN1) (recall footnote in Section 7.2.3.1). This would mean that a child who was tested second in Arabic would have had 3 preceding experiences with the MAIN task before telling Arabic MAIN2 (i.e. telling the Swedish MAIN1 and MAIN2 in the first session, and telling Arabic MAIN1 in the second session). This might have given the child enough training experiences to score high on Arabic MAIN2.

The English-Swedish-speaking bilingual children in Bohnacker’s (2016) study performed similarly in their two languages on MAIN2 (MAIN1 was not elicited). Lindgren & Bohnacker (2022) reported that the German-Swedish-speaking children in their study performed significantly better in Swedish in both MAIN1 and MAIN2. Bohnacker et al. (2022) found no difference in performance between the languages of 100 Turkish-Swedish-speaking children (4–7) but they did find a testing order effect on Turkish scores where children scored higher if tested in Turkish second (and in Swedish first). The authors speculate that the children are exposed to more narrative activities in Swedish at (pre)school, whereas literary language input in Turkish is usually more dependent on language activities at home which might vary from one child to another.

The Arabic-Swedish-speaking children’s narrative production scores did not correlate as strongly between the languages, unlike for narrative comprehension. For production, the correlations were either weak or non-existent (recall Section 7.2.3.1). This means that children who scored high on narrative production in one language did not necessarily score high on narrative production in the other language (in line with Lindgren and Bohnacker, 2022). Furthermore, children scored significantly higher in Arabic MAIN2 than Swedish MAIN2. These findings go against a potential wholesale transfer of macrostructure skills between languages (universality of macrostructure hypothesis) that claims that bilingual children perform similarly with regard to macrostructure in their two languages (e.g. Paradis et al., 2011; Pearson, 2002). Additionally, when observing examples of individual children who performed poorly on narrative production in comparison to their same-aged peers (recall Section 7.2.3.2 and Section 7.2.3.3), many of the low-scoring

---

200 Note that, the 6-year-olds did not score significantly higher on Arabic vocabulary than on Swedish vocabulary, nor did they score significantly higher on Arabic MAIN1 than on Swedish MAIN1.
children in narrative production shared common features of low vocabulary (CLT) scores, low current daily language exposure and/or a relative late AoO of the respective language. Not having sufficient vocabulary knowledge (in addition to basic grammar skills) in a language makes it difficult to express oneself in a narrative production task in that language, which presumably results in low narrative production scores. Future analyses of the present data should statistically test whether these background factors (and others) have a significant effect on the Arabic-Swedish-speaking children’s narrative production scores in each of their languages.

7.3.2.3 Task and story effects

Unlike for narrative comprehension, for narrative production there was no task effect for MAIN1 vs. MAIN2. The Arabic-Swedish-speaking children did not score significantly higher on the narrative production of MAIN1 than on MAIN2. This result is expected since the stories were constructed to be parallel in terms of macrostructure and differed only in the number of characters and their types, see Gagarina et al. (2012). Furthermore, MAIN1 and MAIN2 scores in each language correlated with each other.

No task effects were found either by Bohnacker et al. (2022) for Turkish-Swedish-speaking children and Lindgren (2018) for Swedish-speaking monolinguals, German-Swedish-speaking children and Turkish-Swedish-speaking children. For narrative production, the difficulty level of MAIN1 and MAIN2 in production appears to be equivalent, as opposed to in narrative comprehension. Öztekin (2019, p. 191) attributes this lack of a significant difference between MAIN1 and MAIN2 in her Turkish-Swedish data to the generally low scores in all age groups (most of the children scored less than 50%) that are not high enough to reflect any existing difficulty between the two tasks. It is possible that differences between MAIN1 and MAIN2 in narrative production would emerge if children older than those participating in the present study (4–7) participated.

Furthermore, no story effects were found. The Arabic-Swedish-speaking children scored equally well on the stories of MAIN1 (Cat and Dog), and of MAIN2 (Baby Birds and Baby Goats). These results echo the children’s narrative comprehension results where no story effects were found. Bohnacker et al. (2022) did not find any story effects either for the Turkish-Swedish-speaking children.

The Arabic-Swedish-speaking children had in general more difficulty providing the correct label for the ‘goat’ in Baby Goats story, however, they were quite creative in finding alternative labels (e.g. cows, horses, giraffes). As long as the reference to whom the ‘new label’ referred was clear and consistent, the child’s narrative production score was not affected by the character’s label.
7.3.2.4 Production of story components and story complexity

For both narrative tasks in both languages, the macrostructure components that were expressed with the highest provision rates were attempts and outcomes for all age groups. This is in line with several studies that have used MAIN (Kapalková et al. (2016): Slovak-English-speaking bilinguals; Bohnacker (2016): English-Swedish-speaking bilinguals; Lindgren (2018): Swedish-speaking monolinguals, Turkish-Swedish-speaking bilinguals, German-Swedish-speaking bilinguals, and Öztekin (2019): Turkish-Swedish-speaking bilinguals).

In both Arabic and Swedish, outcomes had higher provision rates in MAIN1 than in MAIN2. Furthermore, age development for attempt and outcome for to be steeper than for other macrostructure components, more so in Swedish than in Arabic. These results are in line with Öztekin (2019) and Fiani et al. (2022) who each found that goals were produced to a lesser degree than attempts and outcomes, which in their turn, both increased with age.

The production of goals was not as frequent as the comprehension of goals, which is in line with previous studies (e.g. Stein & Glenn, 1979, Trabasso et al., 1992) and previous studies using MAIN (e.g. Bohnacker, 2016; Kunnari et al., 2016; Lindgren, 2018; Öztekin, 2019; Lindgren & Bohnacker, 2020). In both languages, the improvement with age for goals was more visible in MAIN2 than in MAIN1. Goal production in Arabic MAIN1 surprisingly decreased with age while goal production in Swedish MAIN1 increased from the 4-year-olds to the 5-year-old group, but then remained constant for the 6-year-old and the 7-year-old groups. These rather odd patterns for the production of goals in Arabic MAIN1 could be a result of sampling in the cross-sectional study, and do not necessarily represent the longitudinal age development of goal production in MAIN1. On the other hand, Bohnacker (2016) found no significant increase in the production of goals over time from age 5 to 7 either, however, her study included only MAIN2 (Baby Birds/Baby Goats). Whether there is a statistically significant increase in the production of goals in the present study remains to be explored.

Internal state terms as initiating events and internal state terms as reaction were very infrequently produced in both languages and both narration tasks. Internal state terms as reaction were the least produced component (among all of the macrostructure components), whereas internal state terms as initiating events were produced to a higher degree with signs of development with age, especially in Swedish. Similarly to goals, internal state terms as initiating events had much lower provision rates in story telling than the understanding of them (as probed by the comprehension questions). This is in line with previous studies (Bohnacker, 2016; Lindgren, 2018; Stein & Glenn, 1979; Trabasso et al., 1992; Öztekin, 2019) that also found observable differences between the children’s production of internal state terms and the comprehension of them.
The finding of the present study that visible components (attempts, outcomes) are produced more frequently than inferred components (goals, internal state terms) confirms Berman and Slobin’s (1994, p.64) proposal that ‘for young children, the salience of individual pictures scenes is what counts rather than a structurally motivated hierarchy of narrative importance.’

As for the production of setting (time and place), in general, there were lower provision rates in MAIN1 than in MAIN2. However, both narration tasks showed a general improvement with age. Expressing setting was more frequent for MAIN2 than for MAIN1, probably due to the fact that the location in MAIN2 plays an essential part in the story. This holds especially true for the Baby Goats story where the baby goat has fallen into the water (of a lake) which is scored as a correct setting if mentioned very early in the narration (according to the BiLI-TAS scoring guidelines, Bohnacker, 2018b). For MAIN1 (Cat/Dog), elements related to place (the bush/the tree) are more frequently mentioned in relation to the outcome of the first episode (the cat fell into the bush/the dog bumped his head on the tree) rather than as an introductory setting.

These results are in line with Lindgren (2018) and Öztekin (2019) who also found that setting was produced more in MAIN2 than in MAIN1. Bohnacker (2016) found a clear age development for settings in MAIN2 (from age 5 to age 6-7). Bohnacker also mentions that many of the older children tended to start their narration with stereotypical fairytale openings (e.g. Once upon a time...). This is in line with results from the present study where the rates for setting at age 7 were much higher than those of the younger age groups, especially in the majority language Swedish (e.g. Swedish Baby Goats by BiAra7-03, det var en gång (..) tre (..) &m getar [: getter] (...) ett (..) lamm (..) &eh drunknade i &eh (..) &eh (.) vattnet ‘once upon a time (..) three (..) &m goates [: goats] (...) one (..) sheep (..) &eh drowned in &eh (..) &eh (.) the water’.

One explanation for why the 7-year-olds mention more settings could be because older children are more aware of the non-shared visuals between themselves and the experimenter than younger children, and as a result provide more information regarding context of time and place for the ‘unknowledgeable’ listener. In other words, older children have more developed cognitive abilities and theory of mind. Another explanation could be related to the older children’s enrolment in school where they may be in more contact with stories and narrative tasks than the younger children at preschool, i.e. more exposed to literacy and literacy conventions. Interestingly, at age 7, the Swedish provision rates for setting in MAIN1 and MAIN2 were much higher than those for Arabic. This is probably related to the schooling language, Swedish, where children read and listen to stories on a daily basis.

In general, although not analysed statistically, improvement with age for narrative macrostructure components was more visible in MAIN2 than in MAIN1, and more so in Swedish than in Arabic. From age 4 to 7, children
seem to develop the ability to include components of the visible actions (attempts and outcomes) to a greater degree than the ability to mention components that need to be inferred from the pictures (goals and internal states).

As for story complexity, provision rates developed in a fairly similar manner in three of the four narration tasks. The children’s story complexity in Arabic MAIN1 did not follow the same pattern as that of Arabic MAIN2, Swedish MAIN1 and Swedish MAIN2. In fact, there were no clear age development patterns in any of the story complexity sequences in Arabic MAIN1. The only visible age development was the decrease of goal-attempt/goal-outcome (GA/GO). The most frequently produced complexity type was ‘no sequence’ in all age groups. Unexpectedly, the production of full episodes (GAOs) in Arabic MAIN1 in the youngest two age groups was more frequent than in the Arabic MAIN2 and the Swedish narration tasks.

Arabic MAIN2, Swedish MAIN1, and Swedish MAIN2 showed a similar pattern in the production of story complexity. The production of ‘no sequences’ decreased with age while the production of complete episodes (GAOs) and attempt-outcomes (AO) gradually increased. However, similarly to Arabic MAIN1, the most frequently produced complexity type was ‘no sequence’. Despite the decrease with age for ‘no sequence’ in Arabic MAIN2 and Swedish MAIN1 and MAIN2, the production of ‘no sequence’ was still quite high at age 7 (between 40.3%–45.8%). These findings are in line with the Turkish-Swedish-speaking children in Öztekin (2019) for whom the rate of ‘no sequence’ at age 7 was still high (between 34.5%–55.9%). Both Lindgren (2018) and Öztekin (2019) found a clear increase with age in the production of full episodes, similar to the results of the present study (with the exception of Arabic MAIN1). This indicates that, in general, older children produce narratives with a higher level of macrostructural complexity and are better in producing sequences of macrostructural components than younger children (also in line with Bohnacker, 2016; Lindgren & Bohnacker, 2022; Fiani et al., 2022). Nevertheless, the production of full episodes (GAOs) was still relatively rare in all age groups (below 20%). By contrast, Gagarina et al., (2019) found that adults produced full episodes (GAO) 40% of the time, and only 7%–10% ‘no sequence’. Moreover, ca. 77% of the adults produced at least one GAO per story. Thus, the results in the current study indicate that narratives produced by Arabic-Swedish-speaking children at age 7 are far from adult-like.

The rather odd results obtained for Arabic MAIN1 could not be explained by effects of testing order. Recall that the children’s narrative production results were very scattered with a wide score range in all age groups. According to the existing literature on other languages, the relatively high production of full episodes (GAOs) and goal-attempt/goal-outcome (GA/GO) is unusual at a young age, since goals are not very frequently expressed at 4 or 5 years of age, in contrast to ‘no sequence’ which is far more common for these age groups (e.g. Bohnacker, 2016; Lindgren, 2018; Öztekin, 2019; Fiani et al.,
It is common for 3–4-year-olds to describe (identify and name) individual objects, characters and events depicted in narrative elicitation materials rather than establishing narrative sequences that have an overall plotline (Berman, 1988; Berman & Slobin, 1994; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994). So, what then makes Arabic MAIN1 an easy task for the children in the current study? The younger children had Arabic as their strongest language, hence, one would expect them to perform just as well on Arabic MAIN2 as on Arabic MAIN1, but they did not. Recall that the 4-year-olds did score significantly higher on Arabic MAIN1 and MAIN2 than on Swedish MAIN1 and MAIN2, respectively. However, the 4-year-olds did not score as well as the older two age groups in Arabic MAIN2 as on MAIN1. Could it be because MAIN1 is easier than MAIN2 in narrative production? Had that been true, then the younger children should have performed just as well in Swedish MAIN1 as they did on Arabic MAIN1. The mystery behind the high performance of the younger children on Arabic MAIN1 remains open for future research.

7.3.3 Concluding remarks

In general and as expected, the children scored higher on narrative comprehension than on narrative production. The results of the present study are in line with earlier work showing that narrative comprehension develops earlier than narrative production and that younger children understand far more than they narrate (e.g. Stein & Glenn, 1979; Trabasso et al., 1992; Bohnacker, 2016; Kapalková et al., 2016; Lindgren, 2018). However, in both languages, the narrative comprehension and narrative production scores of each task correlated with each other indicating that in order for a child to tell a story, s/he must have an understanding of the story elements and how they are related to each other (e.g. Stein & Glenn, 1979; Trabasso & Nickels, 1992; Westby, 2012).

The comprehension of goals and internal state terms was much more accurate than the production of these components in story telling. In narrative comprehension, there were task effects (MAIN1 scores were higher than MAIN2 scores in both languages) whereas no differences between the narrative tasks were found in narrative production. Furthermore, training effects were found in narrative comprehension for Swedish, where children tested in Swedish second benefited from having been tested in Arabic first. No training effects were found for narrative production for the whole sample, however, when broken down by age group, some testing order effects were found. Concerning the children’s performance across their two languages, the children performed similarly in narrative comprehension in both MAIN1 and MAIN2. Narrative production results were different, where the children scored similarly across the languages in MAIN1 but not in MAIN2 (Arabic scores were significantly higher for all age groups except for the 7-year-olds.
where differences stopped to exist). The most frequently produced sequence was ‘no sequence’ whereas the most frequently produced narrative components were attempts and outcomes.

Goals were expressed more frequently in the stories told by the older age groups which meant that the older children were, in general, more able to produce full episodes of narrative complexity (goal-attempt-outcome) than younger children.

At group level, there was an age development for both narrative comprehension and narrative production in both languages. An exception is the unexpected lack of age development for the narrative production of Arabic MAIN1 where more 4-year-olds and 5-year-olds scored relatively high. For narrative comprehension, age development was stronger for Arabic, whereas for narrative production, score development with age was stronger in Swedish. This is not in line with the children’s vocabulary (CLT) results where age development was stronger for Swedish than for Arabic in both vocabulary comprehension and production. For both vocabulary and narrative macrostructure, however, there were great individual variations among the children. Examples of low and high performing children showed reoccurring background factors that may have potentially affected the children’s performance on these language assessment tasks.

For the acquisition of narrative skills, the present study generally shows that by age 7, children’s ability to express attempts and outcomes is fairly established, whereas their abilities to express the inferred narrative components goals and internal states are still under development although they are well understood at this age.
8. Longitudinal study in Sweden: Vocabulary and narrative macrostructure

This chapter starts with stating the specific research questions for the longitudinal study in Sweden (8.1). Then, the children participating in the longitudinal study are presented in relation to their background factors (8.2), followed by the vocabulary results (8.3) and narrative macrostructure results (8.4).

8.1 Research questions

The following research questions are addressed:

- How do comprehension and production of vocabulary (CLT) develop at group level and at individual level from age 4 to age 6?
- How do comprehension and production of narrative macrostructure (MAIN) develop at group level and at individual level from age 4 to age 6?
- Do the longitudinal results match the patterns observed in the larger cross-sectional study?

8.2 Background of the children

Ten 4-year-old children from the cross-sectional study in Sweden participated approximately two years later in the follow-up. As already pointed out in Chapter 5, these 10 children were representative of the larger group in many ways, but not all.

The children in the longitudinal study did not fully represent the cross-sectional sample in relation to place of birth (the proportion of Sweden-born children was greater in the longitudinal sample) and age of onset of Swedish (the proportion of AoO before 3 was greater in the longitudinal sample).

All children in the longitudinal study attended MTI, whether organized via the municipality, a private initiative or both. This differed somewhat from participation in MTI in the larger study, where about two thirds of the children attended MTI. Note, however, that the 6- and 7-year-olds attended more frequently. As for the 6-year-old group in the cross-sectional study, about three-fourths of the children attended MTI.
The participants in the longitudinal study resemble the cross-sectional sample with regard to the proportion of Arabic varieties mostly spoken (Syrian and Palestinian), AoO of Arabic (from birth) and parental SES (similar mean level of parents’ education). Furthermore, in both samples, the parents were reported to speak Arabic with each other and no parent was reported to speak only Swedish to the child.

All parents in the longitudinal study stated that they would advise parents of Arabic-speaking bilingual children to speak the home language at home. Furthermore, all the parents wanted their child to learn Arabic. Some parents were strict regarding issues of code-switching at home, while others had a more lenient attitude and did not insist on the child speaking merely Arabic with them. As for the child’s language use with siblings, a little more than half of the children spoke only Arabic while the rest spoke both Arabic and Swedish. Unlike in the cross-sectional sample, no child in the longitudinal study was reported to speak only Swedish with siblings. However, tendencies towards speaking predominantly Swedish (and only some Arabic) with the siblings were revealed in the parental interviews. Similarly, children’s interest in speaking English had increased by age 6. When speaking to older non-Swedish-speaking relatives, however, the majority of the children were reported to speak Arabic. A detailed description of the children participating in the longitudinal study can be found in Section 5.2.

Despite the small sample size of the longitudinal study ($N = 10$), some statistical analyses were made to explore development in the children’s vocabulary and narrative macrostructure performance from age 4 to age 6.

### 8.3 Results: Vocabulary

In this section, the vocabulary (CLT) results from the children participating in the longitudinal study are analysed. First, an overview of the vocabulary results for both languages (8.3.1) is presented. Then, a comparison between the score development of the children in the smaller and the larger sample is made between ages 4 and 6 (8.3.2). Furthermore, analyses of the development of the vocabulary comprehension (8.3.3) and the vocabulary production (8.3.4) results are presented for both languages. The section continues with an analysis of the performance of individual children (8.3.5), a summary of the results (8.3.6) and a discussion of the findings mainly in relation to the results obtained in the cross-sectional study (8.3.7).

#### 8.3.1 Overview of the Arabic and Swedish vocabulary results

In Table 8.1, the results for vocabulary (CLT) comprehension and production scores (mean, SD and range) are shown for both Arabic and Swedish at Time
1 (age 4) and Time 2 (age 6). At group level, both comprehension and production scores increased from Time 1 to Time 2 in both languages and their ranges narrowed. However, the increase in the mean scores was not the same for the two languages. For Arabic, comprehension scores increased by 10.4 points and production scores increased by 6.6 points. For Swedish, comprehension scores increased by 14.1 points and production by 14.9 points. At Time 1 (age 4), both Arabic comprehension and production scores were higher than the respective Swedish scores. Two years later, at Time 2, the Swedish mean scores had surpassed the Arabic scores in both comprehension and production. At both testing occasions, no child reached ceiling scores (60 points) for any language measure.

Table 8.1. Vocabulary (CLT) scores for Arabic and Swedish at Time 1 (age 4) and Time 2 (age 6) for all follow-up children combined (N = 10).

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Swedish</th>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
<td>Time 1</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>40.0</td>
<td>50.4</td>
<td>37.0</td>
</tr>
<tr>
<td>SD</td>
<td>10.4</td>
<td>6.9</td>
<td>6.4</td>
</tr>
<tr>
<td>Range</td>
<td>25–52</td>
<td>36–58</td>
<td>28–47</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25.7</td>
<td>32.3</td>
<td>21.9</td>
</tr>
<tr>
<td>SD</td>
<td>15.5</td>
<td>18.3</td>
<td>7.0</td>
</tr>
<tr>
<td>Range</td>
<td>1–42</td>
<td>6–52</td>
<td>10–35</td>
</tr>
</tbody>
</table>

Note. For each language, max comprehension score = 60 points, max production score = 60 points.

Unsurprisingly, paired samples t-tests showed that both during Time 1 and Time 2, and in both languages, comprehension scores were significantly higher than production scores (Time 1, Arabic: $t(9) = 6.270, p < .001, d = 1.983$, large effect size; Swedish: $t(9) = 10.261, p < .001, d = 3.245$, large effect size; Time 2, Arabic: $t(9) = 4.794, p = .001, d = 1.516$, large effect size; Swedish: $t(9) = 13.426, p < .001, d = 4.246$, large effect size).

The variation in the scores (SD) was higher in Arabic than in Swedish for both comprehension and production during Time 1 and Time 2. Paired samples t-tests were run to see whether the increase in the scores between Time 1 and Time 2 was significant. The results showed that there was a significant increase in both languages and for both vocabulary measures, all having large effect sizes (Arabic comprehension $t(9) = -5.256, p = .001, d = -1.662$, large effect size; Arabic production $t(9) = -3.808, p = .004, d = -1.204$, large effect size; Swedish comprehension $t(9) = -15.254, p < .001, d = -4.824$, large effect size; Swedish production $t(9) = -8.996, p < .001, d = -2.845$, large effect size).

Although the mean scores for Arabic were higher than Swedish at Time 1, and the mean scores for Swedish were higher than the Arabic scores at Time 2, paired samples t-tests revealed that these differences were not significant.
(Time 1, comprehension: \( t(9) = .910, p = .387; \) production: \( t(9) = .661, p = .525; \) Time 2, comprehension: \( t(9) = -.268, p = .794; \) production: \( t(9) = -.729, p = .485 \)).

8.3.2 Comparing the longitudinal study with the cross-sectional study

This section explores descriptively whether the 10 children in the longitudinal study scored comparably to the 4-year-old group in the cross-sectional study, and whether these 10 children’s vocabulary results at age 6 (Time 2) match the 6-year-old group in the cross-sectional study.

As can be seen in Table 8.2, the longitudinal children mirror the cross-sectional children in terms of the increase in scores between ages 4 and 6 in both Arabic comprehension and Arabic production.

**Table 8.2.** Arabic vocabulary (CLT) at ages 4 and 6 in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study</th>
<th></th>
<th>Longitudinal study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds (N = 22)</td>
<td>6-year-olds (N = 29)</td>
<td>4-year-olds (N = 10)</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>41.6</td>
<td>48.5</td>
<td>40.0</td>
</tr>
<tr>
<td>SD</td>
<td>7.6</td>
<td>7.5</td>
<td>10.4</td>
</tr>
<tr>
<td>Range</td>
<td>25–52</td>
<td>31–58</td>
<td>25–52</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25.6</td>
<td>34.5</td>
<td>25.7</td>
</tr>
<tr>
<td>SD</td>
<td>12.2</td>
<td>13.2</td>
<td>15.5</td>
</tr>
<tr>
<td>Range</td>
<td>1–42</td>
<td>10–53</td>
<td>1–42</td>
</tr>
</tbody>
</table>

*Note.* For each language, max comprehension score = 60 points, max production score = 60 points.

**Table 8.3.** Swedish vocabulary (CLT) at ages 4 and 6 in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study</th>
<th></th>
<th>Longitudinal study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds (N = 22)</td>
<td>6-year-olds (N = 29)</td>
<td>4-year-olds (N = 10)</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>36.1</td>
<td>46.9</td>
<td>37.0</td>
</tr>
<tr>
<td>SD</td>
<td>8.3</td>
<td>10.2</td>
<td>6.4</td>
</tr>
<tr>
<td>Range</td>
<td>18–52</td>
<td>27–60</td>
<td>28–47</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>22.3</td>
<td>31.6</td>
<td>21.9</td>
</tr>
<tr>
<td>SD</td>
<td>7.2</td>
<td>11.5</td>
<td>7.0</td>
</tr>
</tbody>
</table>

*Note.* For each language, max comprehension score = 60 points, max production score = 60 points.
Table 8.3 compares the Swedish comprehension and production scores at ages 4 and 6 in the cross-sectional and longitudinal studies. Similarly to the Arabic scores, the increase in Swedish comprehension and production scores from age 4 to 6 in the longitudinal study mirrors that of the cross-sectional study.

8.3.3 Development of vocabulary comprehension

Figure 8.1 shows boxplots of comprehension scores for Arabic and Swedish (at group level, \(N = 10\)) at Time 1 and Time 2.

As can be seen in Figure 8.1, for both languages, the score variation (SD) shrunk between Time 1 and Time 2. The decrease was, however, more extreme for Arabic than for Swedish. As previously mentioned, there was no significant difference between the comprehension scores in the two languages at Time 1 or Time 2.

In what follows, a closer analysis is provided for each child’s individual comprehension scores between Time 1 (age 4) and Time 2 (age 6). Starting with Arabic, Figure 8.2 shows the scores of each child at both testing times in Arabic comprehension. The distance between the circle (Time 1) and the square (Time 2) shows the development of each child.

As visible in Figure 8.2, all children scored higher at Time 2 than Time 1 in Arabic comprehension. The three children (BiAra4-02, BiAra4-15 and BiAra4-16) who scored below 50% (30 points) at age 4 (Time 1) performed better at age 6 (Time 2); however they were still scoring lowest amongst their peers two years later. BiAra4-02’s score increased by 7 points whereas
BiAra4-15 and BiAra4-16 increased their scores by 21 and 22 points, respectively. The latter two children had the highest increase in scores. The remaining children’s change in score was between 5 and 12 points. Even though no child fully reached ceiling at Time 2, five children (BiAra4-05, BiAra4-08, BiAra4-13, BiAra4-18 and BiAra4-24) scored 90% (54 points) or higher.

Figure 8.2. Scatterplot of Arabic vocabulary (CLT) comprehension scores of the individual children at Time 1 and Time 2. Max score = 60 points.

Figure 8.3. shows the scores of each child at Time 1 and Time 2 in Swedish comprehension. Similarly to Arabic, Swedish comprehension scores increased for all children between Time 1 and Time 2.

Figure 8.3. Scatterplot of Swedish vocabulary (CLT) comprehension scores of the individual children at Time 1 and Time 2. Max score = 60 points.
As can be seen in Figure 8.3, two children (BiAra4-08 and BiAra4-16) had scored just below 50% at age 4, but their scores increased by 15 and 16 points, respectively, by age 6. In fact, most children’s Swedish comprehension scores increased by 15 to 18 points, except for BiAra4-06 whose scores only increased by 8 points, while BiAra4-14’s score increased by 11 points and BiAra4-05 and BiAra4-15’s score increased by 13 points. Similarly to the Arabic comprehension scores, no child managed to score at ceiling (60 points) in Swedish comprehension at Time 2, but four children (BiAra4-06, BiAra5-13, BiAra5-14 and BiAra24) scored above 90% (54 points).

8.3.4 Development of vocabulary production

Figure 8.4 shows boxplots of production scores for Arabic and Swedish (at group level, N = 10) at Time 1 and Time 2.

As mentioned earlier, no significant difference was found between the production scores of the two languages at Time 1 or Time 2. Unlike for comprehension, the score variation for Arabic production increased from age 4 to age 6 by nearly 2.8 SDs, while the Swedish variation decreased by 0.7 SD. This suggests that the children had more uniform scores in Swedish production both at Time 1 and Time 2, as can clearly be seen in Figure 8.4.

Below is a description of each child’s individual score in both Arabic and Swedish production between Time 1 (age 4) and Time 2 (age 6). Figure 8.5 shows the scores at both testing times in Arabic production. The distance between the circle (Time 1) and the square (Time 2) shows the development of each child.
Figure 8.5. Scatterplot of Arabic vocabulary (CLT) production scores of the individual children at Time 1 and Time 2. Max score = 60 points.

While Arabic mean production at group level increased by 6.6 points (recall Table 8.2), a similar increase did not occur for every child. In fact, two children (BiAra4-06 and BiAra4-14) scored 2 points lower at Time 2 than at Time 1. For the remaining children, the increase in scores ranged between 4 points (BiAra4-05) and 15 points (BiAra4-18). While no child scored under 50% (30 points) at Time 2 in Arabic comprehension, the five children who scored below 50% at Time 1 still scored under 30 points at Time 2 in Arabic production (BiAra4-02, BiAra4-06, BiAra4-14, BiAra4-15, BiAra4-16). No child scored 90% (54 points) or above at Time 2. The highest performing children scored 51 and 52 points (BiAra4-08 and BiAra4-18, respectively).

Figure 8.6 shows the scores of each child at Time 1 and Time 2 in Swedish production. Unlike the Arabic production scores, Swedish production scores increased for all children between age 4 (Time 1) and age 6 (Time 2).

Only one child (BiAra4-16) still scored below 50% (30 points) at Time 2. This child had the smallest increase (6 points) between the two test sessions. As for the remaining children, the increase in scores ranged between 11 points (BiAra4-06) and an impressive 24 points (BiAra4-13). Similarly to Arabic production, no child reached 90% (54 points) or higher in Swedish production at Time 2. In fact, the highest score was merely 46 points (77%) by BiAra4-06, who had already scored highest in Swedish production amongst her peers at age 4 (Time 1).
8.3.5 Individual children’s vocabulary scores and background factors

In this section, the performance of individual children on the vocabulary (CLT) comprehension and production tasks will be discussed in connection with certain background factors, starting with Arabic.

In Arabic, all the children’s comprehension scores increased by Time 2 (age 6), yet no child scored at fully ceiling. The majority of the children’s scores increased between 5 and 12 points, however, two children, BiAra4-15 and BiAra4-16, stuck out with an increase of 21 and 22 points, respectively. Both of these children’s parents indicated that their communication with the child at home was almost only/mostly in Arabic. Despite BiAra4-15 and BiAra4-16’s large increase in Arabic comprehension scores, both children still scored relatively low at Time 2, yet above 50% (scoring at 46 and 48 points, respectively). Both children were reported to have one parent speak both Arabic and Swedish while the other parent spoke mostly/only Arabic with them. The parents of BiAra4-06 mentioned in the parental interview that they had started to speak more Swedish to their child since they had experienced that the child had difficulty in understanding everything that was spoken to her in Arabic.

Although BiAra4-02 scored slightly higher than BiAra4-15 and BiAra4-16 (but just below 50%) in Arabic comprehension at Time 1, BiAra4-02’s increase in scores between Time 1 and Time 2 was only 7 points. During the

---

201 Recall that the highest score reached was 58 points (BiAra4-24) followed by 56 points (BiAra4-18) out of a max 60 points.
parental interview it became evident that the parents often code-switch between Arabic and Swedish at home. The child replies to the parents in Swedish and often asks for a Swedish clarification or translation when he does not understand what he is told in Arabic. The child was reported to have an even language exposure (50% Arabic–50% Swedish) throughout the day.

Finally, even though BiAra4-14’s Arabic comprehension scores had increased 10 points since Time 1, the child still scored relatively low (second to lowest) at Time 2. The parents reported that the child ‘understood Arabic when spoken to by the parents’ but refused to speak with them in Arabic. The child was reported to have an even language exposure throughout the day. However, the parents reported that one of them spoke to the child only in Arabic, while the other parent spoke both in Arabic and Swedish.

Moving on to the Arabic production scores, five children (BiAra4-02, BiAra4-06 and BiAra4-14, in addition to BiAra4-15 and BiAra4-16) scored below 30 points (50%) at Time 2. When reviewing the parent interview for these children, all of the parents revealed (in one way or another) that their child preferred to speak Swedish over Arabic. Furthermore, all of the five children were reported to hear and speak more and more English (from school and the internet/television). The parents of BiAra4-14 (who scored 9 points at Time 1, regressing to 7 points at Time 2 in Arabic production) reported in the interview that they were aware that their child mostly spoke to them in Swedish and that the only time the child spoke Arabic intensively was when the family visited relatives in an Arabic-speaking country. Despite continuous efforts for a monolingual setting in the Arabic testing session, BiAra4-14 replied several times in Swedish in the Arabic (CLT) production task.

For BiAra4-15, it was revealed in the parental interview that the child preferred to speak Swedish with her siblings and at home in general. If the child was spoken to in Arabic and the child switched to Swedish, BiAra4-15’s parents mentioned that they would not insist that their child switch back to speaking Arabic. In fact, many of the BiAra4-15’s vocabulary responses on the Arabic CLT were in Swedish despite the experimenter’s repeated prompting for the Arabic word. As for BiAra4-16, the child spoke almost only Arabic with his parents and siblings at home and was often encouraged to continue doing so. When looking closer at the Arabic production test session, BiAra4-16’s replies were found to include several ‘I don’t know’, general-purpose words (for instance, in Lebanese Arabic yimšî ‘walk/go’ instead of the more specific target ydabdib ‘crawl’, or ykassir ‘break (into pieces)’ instead of yinšur ‘saw’) and to describe content of the pictures rather than using the target word, which resulted in a low score. Very seldom did BiAra4-16 use Swedish during the Arabic testing session.

Furthermore, in the parental questionnaire, it was reported that one parent spoke both Arabic and Swedish while the other spoke only Arabic.

278
The remaining five children (BiAra4-05, BiAra4-08, BiAra4-13, BiAra4-18 and BiAra4-24) scored above 50% (30 points)\(^{203}\) on Arabic production at Time 2. As per the parental interviews, all of these children spoke almost only Arabic at home with their parents and siblings. All of the children had Arabic-speaking friends, either at school or at activity centres. One child (BiAra4-05), the lowest scorer in Arabic production amongst the five above-mentioned children, had recently started to speak more English at home and with his friends. The parents reported that they encouraged the child to speak English with them and with his sibling. None of the parents of the remaining four children (BiAra4-08, BiAra4-13, BiAra4-18 and BiAra4-24) mentioned that their child spoke English or Swedish with their siblings.

BiAra4-02 scored extremely low in Arabic production both at Time 1 (1 point) and Time 2 (6 points). In the interview, the parents told that their child has preferred speaking Swedish (over Arabic) since an early age. The child spoke only Swedish with his (older) siblings and he had been very interested in learning English in the past few years. When the parents spoke to the child in Arabic, BiAra4-02 would (almost) always reply in Swedish. The parents added that the child had no Arabic-speaking close friends, and would often speak in Swedish with his Arabic-speaking cousins residing in Sweden. Additionally, BiAra4-02 chose to speak English (rather than Arabic) with (non-Swedish-speaking) relatives residing outside of Sweden. BiAra4-02’s scores and score increase looked very different in Swedish. The child’s Swedish comprehension and production scores each increased by 16 and 19 points, respectively, between Time 1 and Time 2.

As for the Swedish vocabulary (CLT) scores, the 10 children in the follow-up study showed a large increase in their Swedish comprehension and production after two years of being tested at Time 1. All children had been frequently exposed to Swedish through preschool. In Swedish comprehension, the children’s scores increased between 8 points (BiAra4-06) and 18 points (BiAra4-24). Child BiAra4-06, whose scores increased the least, scored the highest amongst his peers at Time 1 (47 points) and well above the group mean (51.10) at Time 2 with 55 points. BiAra4-08 and BiAra4-16, who had scored just below 50% (30 points) at Time 1, increased their Swedish comprehension scores by 15 and 16 points, respectively, by Time 2. Despite this large increase in scores, both children scored well below the mean group score in Swedish comprehension. At Time 2, BiAra4-08 scored 43 points and BiAra4-16 scored 45 points in Swedish comprehension, while the group mean score was 51.10.

\(^{203}\) In fact, these children scored between 42 (70%) and 52 (87%) points on Arabic production at Time 2.
Both BiAra4-08 and BiAra4-16 were reported to have the majority of their daily language exposure in Arabic.  

In Swedish production, the children’s increase in scores ranged between 6 points (BiAra4-16) and 24 points (BiAra4-13). BiAra4-16’s Swedish production scores remained below 50% during Time 1 and Time 2. BiAra4-13, who had the sharpest increase in Swedish production points, scored second to lowest amongst his peers at Time 1, but second to highest at Time 2. In the parental interview, this child was reported to have several Arabic-speaking friends at school. At home, the child would speak (almost) only Arabic with one parent, and a mix of Swedish and Arabic with the other parent. The child is described to have a love for languages and often tries to speak some English with her non-Swedish-speaking cousins and watch Spanish-speaking cartoons on the internet. BiAra4-08 had the second to largest increase (20 points) and scored just at 50% at Time 2. BiAra4-06 and BiAra4-14 had a relatively small increase in points between Time 1 and Time 2 (11 and 10 points, respectively); however, these two children scored the highest amongst their peers at both Time 1 and Time 2. In the parental interview, both children were reported to prefer speaking Swedish (over Arabic) with parents and siblings. BiAra4-06 was reported to have the majority (60%) of her daily language exposure to Swedish.

8.3.6 Development of vocabulary: Summary

At group level, vocabulary (CLT) scores in Arabic and Swedish comprehension and production increased significantly from age 4 (Time 1) to age 6 (Time 2). The Swedish vocabulary scores increased more strongly than the Arabic vocabulary scores, however, there was no significant difference between the scores of the two languages at either testing time. The standard deviation for the comprehension scores in both languages decreased from Time 1 to Time 2, indicating that the scores had become more homogeneous as the children grew up. As for Arabic production scores, the standard deviation had increased by Time 2 indicating a larger variation between the children’s performance. For Swedish production scores, the small standard deviation decreased just slightly since Time 1 suggesting that the children’s scores were and remained quite homogeneous.

---

204 For BiAra4-16, the parents had reported a daily language exposure of 80% Arabic–20% Swedish. For BiAra4-08, the parents did not fill in the daily language exposure question, however it was evident from the parental interview that BiAra4-08 had *very much* Arabic input during the day.

205 Recall that BiAra4-16 also scored well below 50% (30 points) in Arabic production at both Time 1 and Time 2.

206 Recall that BiAra4-13 also scored well above the mean in both Arabic comprehension and production at both Time 1 and Time 2.
At the individual level, for Arabic comprehension, all the children’s scores increased to varying degrees. For Arabic production, all children except two increased their scores from Time 1 to Time 2. While all children scored above 50% (30 points) in Arabic comprehension by Time 2, only half of the children did so in Arabic production by age 6. Swedish comprehension and production scores increased for all children (to varying degrees) between Time 1 and Time 2. Only one child did not manage to score 50% or above (in Swedish production) by age 6. Most of the low-scoring children’s vocabulary (CLT) results could be explained by low current daily language exposure patterns and/or language use at home.

8.3.7 Development of vocabulary: Discussion

The development of vocabulary (CLT) scores in the longitudinal study resembled to a high degree the scores of the 4-year-olds and 6-year-olds in the cross-sectional study. The children in the longitudinal study, however, had an earlier AoO of Swedish (and more children were born in Sweden) than the children in the cross-sectional study. This might explain why the mean Arabic production score of the 6-year-old children in the cross-sectional study was higher than the mean score of the 6-year-olds in the longitudinal study, while the 6-year-olds in the longitudinal study had a higher mean score in Swedish production than the 6-year-olds in the cross-sectional study.

In the cross-sectional study, comprehension scores were significantly higher than the production scores across all age groups. There was a significant difference in performance between the 4-year-olds and the 6-year-olds in Arabic comprehension, Swedish comprehension and Swedish production; yet no significant difference between the two groups’ Arabic production scores though this difference nearly reached significance ($p = .051$) (recall Sections 6.3.2.1 and 6.3.2.2). In the longitudinal study, the same pattern can be observed, yet with a clear difference between the Arabic production scores between Time 1 and Time 2 ($p = .004$). As for the difference between the two languages, in the cross-sectional study, the 4-year-olds had significantly higher Arabic comprehension scores than Swedish comprehension scores (recall Section 6.3.2.3). In the longitudinal study, which included a selection of the 4-year-olds from the cross-sectional sample, there was no significant difference between the two languages’ comprehension scores or their production scores at age 4 (Time 1) and age 6 (Time 2).

At group level ($N = 10$), Arabic vocabulary scores increased with 10.4 in comprehension and 6.6 in production from Time 1 to Time 2. At an individual level, the scores varied greatly. Arabic comprehension scores increased for all children, though the increase varied from 5 to 22 points. For Arabic production, not all children had an increase in scores. Two children, BiAra4-06 and BiAra4-14, had a slight decrease (of 2 points) between Time 1 and Time 2 in Arabic production. These two children scored the highest among their peers.
in Swedish production both at Time 1 and Time 2. The majority of the low-scoring children in Arabic vocabulary were reported to have their daily language exposure predominantly in Swedish, or an even language exposure. Sources of daily language exposure to the home language are wide and include not only the parents’ language to the child, but also performing certain language stimulating activities. For example, regarding literacy activities in Arabic, not all parents were interested in borrowing Arabic books from the local library, but many mentioned that their child had access to videos and songs via the internet. Furthermore, all children were reported to attend some form of Arabic MTI, either provided by the municipality, by private initiatives/organizations, or both. The majority of the parents reported that their child had some Arabic-speaking friend or extended family member with whom they spoke Arabic; however, the extent of this varied greatly. The parent of one child (BiAra4-02) emphasized the possibility for the child to speak English (instead of Arabic) with his non-Swedish-speaking relatives. While all parents emphasized the importance of learning Arabic and providing as much Arabic input as possible, not all households kept an Arabic-only speaking policy. As the children grew older, they were more inclined to speak with their siblings in Swedish (or even in English). Swedish vocabulary comprehension and production scores showed an increase for all 10 children by 14.1 points for comprehension and 14.9 points for production. By age 6, almost all children scored above 50% in both language measures in Swedish. This increase is not surprising since all children were attending preschool at Time 1 (and even before) and had been constantly doing so when tested two years later (Time 2, age 6).

Statistical analyses from the cross-sectional data tested the effect of several language input factors on a bilingual child’s different vocabulary measures (see Section 6.3). The effect of language exposure and home language use is also visible in the longitudinal study through the observation of the vocabulary development of the individual children. Thus, the vocabulary development observed in the longitudinal study largely confirms the results obtained in the cross-sectional study. Not only do scores (in general) increase as the children become older, but there is a difference in the development of the children’s two languages (the home language and the majority language), where language exposure plays a significant role in shaping this development, especially in the home language. These results are in line with Öztekin (2019) who, in her longitudinal study of 10 4-year-old children that were tested again at age 6, found that Swedish vocabulary (production) scores increased for all children with age whereas the home language (Turkish) scores increased to varying degrees depending on home language input factors.
8.4 Results: Narrative macrostructure

In this section, the development of narrative comprehension (8.4.1) and narrative production (8.4.2), as measured by MAIN, is analysed for the children in the longitudinal study. In each part, an overview of the results for both languages is presented in addition to the score development of individual children. Then, the production of macrostructural components and story complexity is presented (8.4.3) followed by analyses of the performance of individual children (8.4.4). The chapter ends with a summary (8.4.5) and a discussion of the findings in relation mainly to the results obtained in the cross-sectional study (8.4.6).

8.4.1 Development of narrative comprehension

8.4.1.1 Overview of the narrative comprehension results

Table 8.4 provides the mean scores, standard deviation (SD) and range for narrative comprehension for each task (MAIN1: Cat/Dog; MAIN2: Baby Birds/Baby Goats) in both languages at Time 1 (age 4) and Time 2 (age 6).

<table>
<thead>
<tr>
<th></th>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>MAIN1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.9*</td>
<td>7.8</td>
</tr>
<tr>
<td>SD</td>
<td>3.2</td>
<td>2.6</td>
</tr>
<tr>
<td>Range</td>
<td>1–10</td>
<td>3–10</td>
</tr>
<tr>
<td>MAIN2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.2*</td>
<td>6.1</td>
</tr>
<tr>
<td>SD</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Range</td>
<td>0.9–7</td>
<td>3–9</td>
</tr>
</tbody>
</table>

Note. For each language, max score = 10 points. The asterisk (*) indicates that the mean is based on 9 children who completed the task.

At group level, Table 8.4 shows that there was an increase in scores in both languages for both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) from Time 1 to Time 2. For Arabic, MAIN1 scores increased by 2.9 points while the MAIN2 scores increased by 3.2 points from Time 1 to Time 2. For Swedish, the mean scores increased by 4.8 points for MAIN1 and 4.6 points for MAIN2 from Time 1 to Time 2.

207 Recall that a child received 1 point when answering a comprehension question correctly. In the few cases where the experimenter did not ask the question (due to experimenter error), the child was assigned the mean score on that particular question for her/his age group. As can be seen for Arabic MAIN2 at Time 1, the minimum score is 0.9 points indicating that the child was awarded 0.9 points (which was the mean score for the particular question that the child was not asked). See Section 4.5.2.1 on scoring of narrative comprehension.
At Time 1 (age 4), the Arabic mean scores were higher for Arabic than Swedish for both MAIN1 and MAIN2. However, at Time 2 (age 6), the Swedish mean scores had surpassed the Arabic mean scores for MAIN1 but not for MAIN2, although they were very close. At Time 1, ceiling (10 points) was reached for Arabic MAIN1 and almost (9 points) for Swedish MAIN1. At Time 2, ceiling was reached for both Arabic MAIN1 and Swedish MAIN1 but not for MAIN2 in the respective languages.

As for the range (minimum scores–maximum score), interesting changes occurred from Time 1 to Time 2. For Arabic, the range for MAIN1 decreased by 2 points as the minimum score rose from 1 point to 3 points. Some children were able to score at ceiling in MAIN1 even when they were 4 years old. For MAIN2, the maximum score at Time 1 was 7 points, however it increased by 2 points to reach 9 points by Time 2. Also, the minimum score increased from 0.9 points to 3 points by Time 2. As for Swedish, one noticeable change in the score range occurred for MAIN1. At Time 1, the range was very wide (0–9 points), however at Time 2, the range shrunk drastically (7–10 points). For Swedish MAIN2, another noticeable change occurred as the minimum score increased from 0 points to 3 points and the maximum score doubled from 4 points at Time 1 to 8 points at Time 2. The developments with time are visualized with box plots below.

Paired samples t-tests showed that the increase in scores from Time 1 to Time 2 was significant for both languages and both narrative tasks (Arabic MAIN1: $t(8) = 3.077, p = .015, d = 1.026$, large effect size; Arabic MAIN2: $t(8) = 4.845, p = .001, d = 1.615$, large effect size; Swedish MAIN1: $t(8) = 3.867, p = .005, d = 1.289$, large effect size; Swedish MAIN2: $t(9) = 8.699, p < .001, d = 2.751$, large effect size).

As for the difference in performance between the two languages, paired samples t-tests showed that there was a significant difference between the Arabic and Swedish MAIN2 scores at Time 1 ($t(8) = 3.202, p = .013, d = 1.067$, large effect size) where the children scored significantly higher in Arabic. There was no significant difference between the two languages at Time 2 or for the other narrative task (MAIN1 at Time 1: $t(7) = 1.014, p = .344$; MAIN1 at Time 1: $t(9) = 1.349, p = .210$; MAIN2 at Time 2: $t(9) = .227, p = .825$).

Although the Arabic mean score for MAIN1 was higher than that for MAIN2 at Time 1, paired samples t-tests also showed that there was no significant difference between the two tasks ($t(8) = 1.754, p = .118$). At Time 2, there was a significant difference between MAIN1 and MAIN2 ($t(9) = 3.431, p = .008, d = 1.085$, large effect size). For Swedish, MAIN1 mean scores were significantly higher than MAIN2 mean scores, both at Time 1 ($t(8) = 4.584, p = .004, d = 1.310$, large effect size) and at Time 2 ($t(9) = 5.765, p < .001, d = 1.823$, large effect size).

Next, the development of narrative comprehension scores in each language and task (MAIN1 and MAIN2) is analysed separately and for every child.
8.4.1.2 Development of Arabic and Swedish narrative comprehension
Starting with a visualization of the results at group level, Figure 8.7 shows boxplots of the Arabic MAIN1 and Arabic MAIN2 scores for narrative comprehension at Time 1 and Time 2.

![Boxplot of the scores for Arabic narrative comprehension at Time 1 (age 4) and Time 2 (age 6). Max score = 10 points. The horizontal line inside each box represents the median.]

Next, the individual children’s scores on narrative comprehension on Arabic MAIN1 and Arabic MAIN2 are analysed in further detail. Each child’s score at Time 1 and Time 2 is plotted in Figure 8.8 for MAIN1 and Figure 8.9 for MAIN2. The distance between the circle (Time 1) and the square (Time 2) shows the development of each child.

For Arabic MAIN1 (Figure 8.8), seven out of nine children showed a development between Time 1 and Time 2.\(^\text{208}\) The most drastic increase was for BiAra4-05 with an increase of 9 points. The smallest increase (1 point) was for BiAra4-08 who had already scored 9 points at Time 1. Four children (BiAra4-05, BiAra4-08, BiAra4-13, BiAra4-18) scored at ceiling (10 points) at Time 2. At Time 1, four children scored below 50% (below 5 points); however at Time 2, only 1 child (BiAra4-02) scored below 50%. One child (BiAra4-06) scored 5 points both at Time 1 and Time 2. For one child (BiAra4-24) the scores decreased from Time 1 to Time 2, though this decrease was only by 1 point (from 10 points to 9 points).

\(^{208}\) Recall that one child (BiAra4-02) did not perform Arabic MAIN1 and Arabic MAIN2 at Time 1.
For Arabic MAIN2 (Figure 8.9), almost all children’s scores increased from Time 1 to Time 2. One exception was BiAra4-06 whose scores decreased by 1 point (from 4 points to 3 points). The largest increase in scores occurred for BiAra4-05 with an increase in 6 points, and the smallest increase was for BiAra4-13 and BiAra4-14 with 2 points each. BiAra4-13 scored higher than her peers at both Time 1 and Time 2 with 7 points and 9 points, respectively. Unlike the children’s performance on MAIN1 at age 6 where four children
scored at ceiling, the MAIN2 scores did not reach as high. At Time 1, eight out of the nine children scored below 50% (below 5 points), however at Time 2, only 2 children (BiAra4-06 and BiAra4-14) still scored below 5 points. BiAra4-02, who was not able to perform MAIN2 in Arabic at age 4, scored 5 points at age 6.

Below, the children’s Swedish narrative comprehension scores are discussed. Figure 8.10 shows boxplots of the narrative comprehension scores for Swedish MAIN1 and Swedish MAIN2 (at group level) at Time 1 and Time 2.

![Figure 8.10. Boxplot of the scores for Swedish narrative comprehension at Time 1 (age 4) and Time 2 (age 6). Max score = 10 points. The horizontal line inside each box represents the median.](image)

Next, the individual children’s scores on the Swedish narrative comprehension on MAIN1 and MAIN2 are analysed. Each child’s score at Time 1 and Time 2 is provided in Figure 8.11 for Swedish MAIN1 and Figure 8.12 for Swedish MAIN2. The distance between the circle (Time 1) and the square (Time 2) shows the development of each child.

As can be seen in Figure 8.11, all of the children’s scores increased on Swedish MAIN1 between Time 1 and Time 2 except for BiAra2-24 who scored 9 points (almost ceiling level) on both testings. For one child (BiAra4-05), no data is available for Swedish MAIN1 during Time 1. Five children (BiAra4-05, BiAra4-06, BiAra4-14, BiAra4-15, BiAra4-18) scored 10 points at age 6. The sharpest increase in scores was for BiAra4-15 who scored at floor level (0 points) at age 4 and full points (10 points) at age 6. No child scored below 50% (below 5 points) at Time 2, whereas the majority of the children scored at 50% or less during Time 1. One exception was BiAra4-06...
who scored 9 points at Time 1. Seven children scored 90% or above on Swedish MAIN1 at age 6.

Figure 8.11. Swedish MAIN1 (Cat/Dog) narrative comprehension scores of the individual children at Time 1 (age 4) and Time 2 (age 6). Max score = 10 points.

Figure 8.12. Swedish MAIN2 (Baby Birds/Baby Goats) narrative comprehension scores of the individual children at Time 1 (age 4) and Time 2 (age 6). Max score = 10 points.

As for Swedish MAIN2 (Figure 8.12), all of the children’s comprehension scores increased from Time 1 to Time 2. At the first testing (age 4), all of the children scored below 50%, and several children out of them scored at floor
level with 0 points (BiAra4-02, BiAra4-15 and BiAra4-16). At the second testing (age 6), only three children scored below 50% (BiAra4-13, BiAra4-15, BiAra4-16). No child scored 90% or above; the highest score achieved was 8 points (BiAra4-06, BiAra4-08, BiAra4-14). The sharpest increase in scores was for BiAra4-08 and BiAra4-14 who both scored 1 point at Time 1 and 8 points at Time 2. For information about the performance of the individual children (in relation to certain background factors), see Section 8.4.4.

### 8.4.1.3 Comparing the longitudinal study with the cross-sectional study

In this section, the narrative comprehension scores of the 10 children in the longitudinal study at Time 1 (age 4) and Time 2 (age 6) are compared descriptively to their same aged peers in the cross-sectional study. The scores of both studies are presented in Table 8.5 for Arabic and Table 8.6 for Swedish.

**Table 8.5.** Arabic narrative comprehension (MAIN) at age 4 and 6 in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study</th>
<th>Longitudinal study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds (N = 21)</td>
<td>6-year-olds (N = 29)</td>
</tr>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.6</td>
<td>7.2</td>
</tr>
<tr>
<td>SD</td>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Range</td>
<td>1–10</td>
<td>1–10</td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.9</td>
<td>5.5</td>
</tr>
<tr>
<td>SD</td>
<td>1.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Range</td>
<td>0–7</td>
<td>2–10</td>
</tr>
</tbody>
</table>

*Note.* Max narrative comprehension score = 10 points. BB = Baby Birds, BG = Baby Goats. The asterisk (*) indicates that the mean is based on 9 children who completed the task.

**Table 8.6.** Swedish narrative comprehension (MAIN) at age 4 and 6 in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study</th>
<th>Longitudinal study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds (N = 20/22)†</td>
<td>6-year-olds (N = 29)</td>
</tr>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.4</td>
<td>7.4</td>
</tr>
<tr>
<td>SD</td>
<td>3.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Range</td>
<td>0–10</td>
<td>2–10</td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.3</td>
<td>5.0</td>
</tr>
<tr>
<td>SD</td>
<td>2.2</td>
<td>2.8</td>
</tr>
<tr>
<td>Range</td>
<td>0–8</td>
<td>1–10</td>
</tr>
</tbody>
</table>

*Note.* Max narrative comprehension score = 10 points. BB = Baby Birds, BG = Baby Goats. The asterisk (*) indicates that the mean is based on 9 children who completed the task. † For the 4-year-olds, there were answers from 20 children for MAIN1 and from 22 children for MAIN2.
As can be seen, for both Arabic and Swedish narrative comprehension, the increase in mean scores (from age 4 to 6) of the children in the longitudinal study mirrors the increase in scores in the cross-sectional sample. A noticeable difference between the cross-sectional and longitudinal studies was the children’s performance on Swedish MAIN1 where the mean scores were much higher in the longitudinal study and the range was much narrower than in the cross-sectional study.

8.4.2 Development of narrative production

8.4.2.1 Overview of the narrative production results

In Table 8.7, the mean scores, standard deviation (SD) and range for narrative production is provided for each narrative task (MAIN1: Cat/Dog; MAIN2: Baby Birds/Baby Goats) and each language at Time 1 (age 4) and Time 2 (age 6).

Table 8.7. Follow-up children combined (N = 10) narrative production (MAIN) for Arabic and Swedish at Time 1 (age 4) and Time 2 (age 6).

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
</tr>
<tr>
<td>MAIN1</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.0*</td>
</tr>
<tr>
<td>SD</td>
<td>3.9</td>
</tr>
<tr>
<td>Range</td>
<td>0–9</td>
</tr>
<tr>
<td>MAIN2</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.3*</td>
</tr>
<tr>
<td>SD</td>
<td>2.3</td>
</tr>
<tr>
<td>Range</td>
<td>0–8</td>
</tr>
</tbody>
</table>

Note. For each language, max score = 17 points. The asterisk (*) indicates that the mean is based on 9 children who completed the task.

Table 8.7 indicates that there was an increase in scores for both languages for both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) from Time 1 to Time 2, however this increase was the least sharp for Arabic MAIN1 (with an increase of 0.8 points). For Arabic MAIN2, the mean score increased by 2 points between the two testing occasions. For Swedish, the mean scores increased greatly, by 3.1 points and 3.9 points for MAIN1 and MAIN2, respectively.

For both Arabic and Swedish, the MAIN1 mean scores were higher than the MAIN2 mean scores at Time 1, but at Time 2, the MAIN2 mean scores were somewhat higher than the MAIN1 mean scores. The highest points scored were in Swedish MAIN2 during Time 2 (at age 6) with 10 points. No child scored anywhere near ceiling (17 points) in either MAIN1 or MAIN2 on any testing occasion. In fact, for both Arabic and Swedish, the mean scores were quite low and well below 50% (8.5 points). The SD decreased for Arabic
MAIN1 from Time 1 to Time 2 but increased for Arabic MAIN2, Swedish MAIN1 and Swedish MAIN2 between the two testing occasions.

Paired samples t-tests showed that, at group level, there was no significant difference between the scores of the two languages in neither MAIN1 nor MAIN2 on any testing occasion (MAIN1: Time 1, $t(7) = 1.538, p = .168$, Time 2, $t(8) = 1.155, p = .282$; MAIN2: Time 1, $t(8) = 1.960, p = .086$, Time 2, $t(8) = .084, p = .935$).

Similarly, for both Arabic and Swedish, the children did not score significantly higher on MAIN1 than on MAIN2 (Arabic: Time 1, $t(8) = .625, p = .550$, Time 2, $t(8) = .958, p = .366$; Swedish: Time 1 $t(8) = .1033, p = .332$, Time 2, $t(9) = .085, p = .934$). In other words, at group level, the children scored equally well on both tasks (MAIN1 and MAIN2) in both languages at any age.

Comparing the children’s performance between Time 1 and Time 2 in the same language, paired samples t-tests showed that for Arabic MAIN1, there was no significant difference between the children’s performance over time ($t(8) = .631, p = .546$). For Arabic MAIN2, the difference just reached significance ($t(8) = 2.353, p = .046, d = .784$, moderate to large effect size). For Swedish, the difference between Time 1 and Time 2 was significant (with large effect size) for both MAIN1 ($t(8) = 3.054, p = .016, d = 1.081$, large effect size) and for MAIN2 ($t(9) = 4.593, p = .001, d = 1.452$, large effect size) indicating a stronger age development for Swedish than for Arabic.

### 8.4.2.2 Development of Arabic and Swedish production of narrative macrostructure

Figure 8.13 shows the Arabic MAIN1 and Arabic MAIN2 scores for production of narrative macrostructure (at group level) at Time 1 and Time 2.

Although the SD narrowed greatly from Time 1 to Time 2 for Arabic MAIN1 scores, there was no significant difference (at group level) between the two mean scores. For MAIN2, the SD barely changed between the two testing occasions; however, there was a significant increase in the mean score from Time 1 to Time 2. These results mirror the results obtained for narrative comprehension in the longitudinal study where there was no significant difference between the scores of Arabic MAIN1 between Time 1 and Time 2, but a significant increase for Arabic MAIN2.
Next, the development of each child’s individual score in Arabic production of narrative macrostructure in MAIN1 (Figure 8.14) and MAIN2 (Figure 8.15) is visualized.

Figure 8.14. Arabic MAIN1 (Cat/Dog) narrative production scores of the individual children at Time 1 (age 4) and Time 2 (age 6). Max score = 17 points.

Figure 8.14 shows that, in MAIN1, the Arabic narrative production scores of three children (BiAra4-05, BiAra4-06, BiAra4-16) increased from Time 1 to Time 2. The sharpest increase was for BiAra4-16 who scored 6 points higher
at the second testing. The scores of three children (BiAra4-13, BiAra4-18, BiAra4-24) decreased from Time 1 to Time 2. The steepest decrease was for BiAra4-13 whose score dropped from 9 points to 5 points. Finally, for three children (BiAra4-08, BiAra4-14, BiAra4-15), the Arabic production scores between Time 1 and Time 2 did not change.

Figure 8.15. Arabic MAIN2 (Baby Birds/Baby Goats) narrative production scores of the individual children at Time 1 (age 4) and Time 2 (age 6). Max score = 17 points.

As for MAIN2, Figure 8.15 shows that the majority of the children’s scores increased from Time 1 to Time 2 on narrative production. The sharpest increase was for BiAra4-16 and BiAra4-18 whose scores increased by 5 points. The scores of two children, BiAra4-13 and BiAra4-24, decreased by 2 points and 1 point, respectively. One child, BiAra4-14, scored 0 points on both Arabic MAIN1 and MAIN2 at both age 4 and age 6.

At group level, the scores were quite low. For both MAIN1 and MAIN2, the scores of all children at Time 1 (age 4) were below 50% (8.5 points), as were the scores for the children in MAIN2 at Time 2 (age 6). For MAIN1 at Time 1, only two children (BiAra4-13, BiAra4-18) scored above 50% with 9 points each, while the remaining children scored below 8.5 points. Two children (BiAra4-05, BiAra4-16) scored at floor level on MAIN1 at age 4, but no child scored 0 points at age 6. For MAIN2, one child (BiAra4-14) scored 0 points at both testing occasions.

Next, the Swedish production scores are presented. Figure 8.16 shows the scores at group level for the children’s performance on Swedish MAIN1 and Swedish MAIN2 at age 4 (Time 1) and at age 6 (Time 2).
As previously mentioned, the children scored significantly higher in both MAIN1 and MAIN2 during the second Swedish testing (age 6) in comparison to the first testing (age 4). Also, for both Swedish MAIN1 and MAIN2, the SD increased with time.

In what follows, the individual children’s scores on the production of narrative macrostructure on Swedish MAIN1 and Swedish MAIN2 are analysed. Figure 8.17 (MAIN1) and Figure 8.18 (MAIN2) show each child’s score at age 4 (Time 1) and age 6 (Time 2). The distance between the circle (Time 1) and the square (Time 2) shows the development of each child.

Figure 8.17 shows that although the majority of the children’s scores increased in Swedish MAIN1 from Time 1 to Time 2, no child scored above 50% (8.5 points) at any of the testing occasions. The highest score achieved was 8 points by four children (BiAra4-05, BiAra4-06, BiAra4-08, BiAra4-18). BiAra4-05 was not able to perform MAIN1 at Time 1 and the task was prematurely terminated due to the child’s difficulty in telling the story and lack of will to cooperate. For one child (BiAra4-24) the Swedish MAIN1 score decreased by 1 point (from 3 points to 2 points).

As for Swedish MAIN2, Figure 8.18 shows quite a similar pattern with the majority of the children scoring higher at age 6 than at age 4. One child (BiAra4-18) scored the same on both testing occasions (2 points). The sharpest increase was for BiAra4-13 who scored 0 points at Time 1 and 9 points (second to highest) at Time 2. Only two children, BiAra4-08 and BiAra4-13, scored above 50% at Time 2 on Swedish production in MAIN2 with 10 points and 9 points, respectively. However, in general, the scores were quite low for
both MAIN1 and MAIN2 and at both testing occasions. For information about the performance of the individual children in relation to certain background factors, see Section 8.4.4.

![Figure 8.17. Swedish MAIN1 (Cat/Dog) narrative production scores of the individual children at Time 1 (age 4) and Time 2 (age 6). Max score = 17 points.](image1)

![Figure 8.18. Swedish MAIN2 (Baby Birds/Baby Goats) narrative production scores of the individual children at Time 1 (age 4) and Time 2 (age 6). Max score = 17 points.](image2)

8.4.2.3 Comparing the longitudinal study with the cross-sectional sample
This section descriptively compares the narrative production scores of the 10 children in the longitudinal study at Time 1 (age 4) and Time 2 (age 6) to their same aged peers in the cross-sectional study. The scores of both studies are presented in Table 8.8 for Arabic and Table 8.9 for Swedish.
Table 8.8. Arabic narrative production (MAIN) at age 4 and 6 in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study</th>
<th>Longitudinal study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds (N = 21)</td>
<td>6-year-olds (N = 29)</td>
</tr>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.6</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td></td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.7</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td></td>
</tr>
</tbody>
</table>

Note. Max narrative production score = 17 points. BB = Baby Birds, BG = Baby Goats. The asterisk (*) indicates that the mean is based on 9 children who completed the task.

Table 8.9. Swedish narrative production (MAIN) at age 4 and 6 in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional study</th>
<th>Longitudinal study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds (N = 20/22)†</td>
<td>6-year-olds (N = 29)</td>
</tr>
<tr>
<td>MAIN1 (Cat/Dog)</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td></td>
</tr>
<tr>
<td>MAIN2 (BB/BG)</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td></td>
</tr>
</tbody>
</table>

Note. Max narrative production score = 17 points. BB = Baby Birds, BG = Baby Goats. The asterisk (*) indicates that the mean is based on 9 children who completed the task. † For the 4-year-olds, there were answers from 20 children for MAIN1 and from 22 children for MAIN2.

As can be seen in Table 8.8 and Table 8.9, for both languages, the longitudinal study mirrors the cross-sectional study in terms of increase in narrative production scores from age 4 to age 6. The mean scores of the 6-year-olds in the cross-sectional study were slightly higher in Arabic but slightly lower in Swedish compared to the children in the longitudinal study. Also, in both studies, there was a sharper increase in scores for the cross-sectional children in Arabic, but for Swedish, the longitudinal children showed a sharper increase in mean scores from age 4 to age 6.

8.4.3 Macrostructural components and story complexity

In this section, a descriptive comparison is provided for the production of narrative components and story complexity in both languages. The results of the
longitudinal study are compared with the performance of the same aged peers (the 4-year-olds and the 6-year-olds) in the cross-sectional study.

8.4.3.1 Macrostructural components
For exact percentages of the provision of macrostructural components in the longitudinal sample and cross-sectional sample, see Table A3.1 and Table A3.2 in Appendix 3.1.

For the Arabic production of narrative components in MAIN1, the 4-year-olds and the 6-year-olds in the cross-sectional study and longitudinal study performed fairly similarly to each other. In both studies (and both age groups), the most frequently produced macrostructural component was the outcome. One noticeable difference concerned internal states as initiating event (IST as IE) where the children in the cross-sectional study produced much higher proportions than those in the longitudinal study. For the production of setting, it was the other way around, where the children in the longitudinal study produced more. Interestingly, the 6-year-olds did not always perform better than the 4-year-olds in both studies. For example, the rate of goals and internal states as reaction (IST as R) was higher in the 4-year-olds than the 6-year-olds.

Figure 8.19. Production of narrative macrostructure components in Arabic MAIN1 at age 4 and age 6 in the cross-sectional study and the longitudinal study. Cross = cross-sectional, Long = longitudinal.

As for MAIN2, the patterns for production of narrative components were similar. The children produced mostly attempts at both ages and in both samples.
In both the cross-sectional and longitudinal samples, the production of outcomes more than doubled by age 6. A noticeable difference was the comparably high production of setting in the 6-year-olds in the longitudinal study, much higher than in the cross-sectional study. The 6-year-olds in both studies produced more components in MAIN2 than the 4-year-olds, with the exception of the production of goals in the longitudinal study which was noticeably low.

For Swedish, the production of macrostructural components showed a similar pattern for the cross-sectional and the longitudinal studies. For MAIN1, the most frequently produced narrative component was outcomes for both age groups and samples. The rate of provision of narrative components was always higher in the 6-year-olds than in the 4-year-olds. Similarly to Arabic MAIN1 and MAIN2, the production of setting in Swedish MAIN1 by the 6-year-olds in the longitudinal sample was remarkably higher than in the remaining groups.

Similarly to MAIN1, the 6-year-olds scored higher than the 4-year-olds on MAIN2 in both samples. The 4-year-olds in the cross-sectional study scored slightly higher than their peers in the longitudinal study. At age 6, the rate of provision of internal state terms and initiating events, attempts and outcomes was more frequent for the longitudinal children. However, recall that the children in the longitudinal study have an earlier AoO to Swedish than the chil-
dren in the cross-sectional study. Interestingly, unlike for Arabic MAIN1, Arabic MAIN2 and Swedish MAIN1, the 6-year-olds in the longitudinal study did not manage to score any point on setting in Swedish MAIN2.

**Figure 8.21.** Production of narrative macrostructure components in Swedish MAIN1 at age 4 and age 6 in the cross-sectional study and the longitudinal study. Cross = cross-sectional, Long = longitudinal.

**Figure 8.22.** Production of narrative macrostructure components in Swedish MAIN2 at age 4 and age 6 in the cross-sectional study and the longitudinal study. Cross = cross-sectional, Long = longitudinal.
8.4.3.2 Story complexity

Figures 8.23 and 8.24 show the percentage of story complexity produced in Arabic MAIN1 and MAIN2, respectively, for the 4-year-olds and the 6-year-olds in the cross-sectional and the longitudinal studies. Figures 8.25 and 8.26 show the percentage of story complexity produced in Swedish MAIN1 and MAIN2, respectively, for the 4-year-olds and the 6-year-olds in the cross-sectional and the longitudinal studies. For exact percentages of production rate of narrative story complexity in the longitudinal sample and cross-sectional sample, see Table A3.3 and Table A3.4 in Appendix A3.2.

For Arabic MAIN1, the children seem to perform similarly in both age groups and in both samples. The most frequently produced story complexity type is ‘no sequence’. Goals do not increase by age 6. Similarly for MAIN2, ‘no sequence’ is most common, however its rate of provision decreases as the children grow older. In the cross-sectional study, a clear increase in goals can be seen but not in the longitudinal study.

For Swedish, a clearer improvement with age is seen in both the cross-sectional and longitudinal samples. Similarly to Arabic, for both MAIN1 and MAIN2, the most commonly produced complexity type is ‘no sequence’, which clearly decreases by age 6. Furthermore, goals and attempt-outcomes (AO) increase noticeably by age 6 in both samples.

Figure 8.23. Percentage (%) of story complexity in Arabic MAIN1 (Cat/Dog) for the 4-year-olds and the 6-year-olds in the cross-sectional and the longitudinal studies. AO = attempt-outcome, GA/GO = goal-attempt/goal-outcome, GAO = goal-attempt-outcome. Cross = cross-sectional, Long = longitudinal.
Figure 8.24. Percentage (%) of story complexity in Arabic MAIN2 (Baby Birds/Baby Goats) for the 4-year-olds and the 6-year-olds in the cross-sectional and the longitudinal studies. AO = attempt-outcome, GA/GO = goal-attempt/goal-outcome, GAO = goal-attempt-outcome. Cross = cross-sectional, Long = longitudinal.

Figure 8.25. Percentage (%) of story complexity in Swedish MAIN1 (Cat/Dog) for the 4-year-olds and the 6-year-olds in the cross-sectional and the longitudinal studies. AO = attempt-outcome, GA/GO = goal-attempt/goal-outcome, GAO = goal-attempt-outcome. Cross = cross-sectional, Long = longitudinal.
Figure 8.26. Percentage (%) of story complexity in Swedish MAIN2 (Baby Birds/Baby Goats) for the 4-year-olds and the 6-year-olds in the cross-sectional and the longitudinal studies. AO = attempt-outcome, GA/GO = goal-attempt/goal-outcome, GAO = goal-attempt-outcome. Cross = cross-sectional, Long = longitudinal.

In general, the children’s performance in the cross-sectional sample and the longitudinal sample resemble each other more in Swedish than in Arabic. Similarly to vocabulary (CLT) production, there was a clearer improvement with age for the children’s narrative macrostructure production in Swedish than in Arabic.

8.4.4 Individual children’s narrative macrostructure development
As was previously mentioned, at group level, the children’s scores improved on all narrative tasks in both Arabic and Swedish (apart from narrative production of Arabic MAIN1) from age 4 to age 6. At the individual level, the children’s narrative development was far from homogeneous.

This section discusses the individual development of each child in narrative comprehension and narrative production (in both languages) in connection with certain background factors and vocabulary (CLT) score development. Recall that, at group level, the children’s vocabulary (CLT) scores had increased over time for both Arabic and Swedish comprehension and production. For detailed information about the children’s vocabulary development, refer to Section 8.3.

BiAra4-02 did not tell the Arabic MAIN narratives, neither at Time 1 nor at Time 2. The child scored low on Arabic vocabulary (CLT) at both age 4
and age 6. BiAra4-02 had an AoO of Swedish at age 2, but was reported to prefer speaking Swedish over Arabic. In the questionnaire filled in for the longitudinal study, the parents revealed that they had gradually come to speak more Swedish with the child at the age of 6 than when the child was younger. Hence, the child’s lack of Arabic language development might be a result of limited exposure and language use. BiAra4-02’s Swedish narrative comprehension and narrative production scores each increased by 3 points (MAIN1) and 5 points (MAIN2). However, in no task did the child reach ceiling. The highest BiAra4-02 scored was 7 points on narrative comprehension of MAIN1 and 7 points on narrative production of MAIN2, both at age 6. For BiAra4-02’s vocabulary (CLT) scores, the child’s Arabic comprehension and production scores were near floor level, even after 2 years. For Swedish, however, the vocabulary scores increased sharply. Furthermore, the child was reported to speak English (rather than Arabic) with siblings and with non-Swedish speaking members of the extended family. BiAra4-02 may not be on the road to becoming an active bilingual due to limited use and exposure of Arabic.

BiAra4-05 had sharp increases in Arabic narrative comprehension in MAIN1 (with 9 points) and MAIN2 (with 5 points). BiAra4-05’s scores also increased in Arabic narrative production of MAIN1 (5 points increase) and MAIN2 (2 points increase). As for Swedish, upon closer inspection of the video recording, BiAra4-05 appeared to be shy at the beginning of the Swedish testing session (at Time 1) and as a result did not tell MAIN1. (No score was awarded.) BiAra4-05 did tell Swedish MAIN2 but scored very low on both narrative comprehension and narrative production with 1 point on each task. At Time 2 (age 6), BiAra4-05 scored at ceiling on Swedish MAIN1 narrative comprehension and second to highest amongst his peers on Swedish MAIN2 narrative comprehension. For Swedish narrative production, BiAra4-05 scored high on MAIN1 (with 8 points, which was the highest score) but only 2 points (the lowest score) on MAIN2. While telling MAIN2, the child used all-purpose verbs instead of specific verbs that clearly expressed the episodic events (for example, *come* instead of *jump up*).\(^{209}\) Looking at BiAra4-05’s vocabulary (CLT) scores, the child’s Arabic scores were higher than the Swedish vocabulary scores at both Time 1 and Time 2. However, at age 6, Swedish vocabulary scores had increased greatly and almost caught up with the Arabic vocabulary scores. At age 4, the child was reported to have the majority (80%) of his language exposure in Arabic since several of his preschool staff and peers were Arabic speakers. By age 6, the child had started to attend förskoleklass (pre-school class) and was reported to have an even

\(^{209}\) Although still comprehensible to the listener, the child conjugated several high frequency words wrong in the present tense (e.g. *drunkar* instead of *drunknar* ‘drown’) and the past tense (e.g. *komde* instead of *kom* ‘come’; *sågde* instead of *såg* ‘saw’) which indicates his language proficiency is lower than what can be expected of a 6-year-old.
amount of exposure to both languages throughout the day. In the parental interview, BiAra4-05’s parent revealed that the child enjoyed speaking English in addition to Swedish with his sibling, but mostly Arabic with the parents in addition to some English with one of the parents. The additional exposure to a third language may have contributed to why the child’s Swedish vocabulary scores did not quite reach the Arabic ones.

BiAra4-06’s Arabic narrative production scores increased by 7 points (MAIN1) and 3 points (MAIN2). Surprisingly, the child’s Arabic narrative comprehension scores did not increase. In fact, the child scored 5 points at both Time 1 and Time 2 for MAIN1 and the child’s score decreased by 1 point on MAIN2 (from 4 points to 3 points). When further inspecting the child’s narrative comprehension responses, BiAra4-06 seems to have a hard time remembering the Arabic equivalent for ‘happy/sad, good/bad’ and as a result skipped answering several of the IST questions (and their follow-up questions). For Swedish, the narrative comprehension scores for MAIN1 were already high at age 4 and reached ceiling (10 points) at age 6. For MAIN2, the Swedish narrative comprehension scores doubled (from 4 points to 8 points). As for Swedish narrative production, scores on MAIN1 increased sharply from 1 point (Time 1) to 8 points (Time 2). For Swedish MAIN2, narrative production scores also doubled (from 3 points at Time 1 to 6 points at Time 2). Looking at the child’s vocabulary (CLT) scores, BiAra4-06’s Arabic scores remained almost unchanged from Time 1 to Time2, whereas the child’s Swedish vocabulary increases greatly. Regarding the child’s language exposure, BiAra4-06 was reported to attend a school where both Swedish and English are used daily for classroom instruction. At home, the child speaks mostly in Swedish with her sibling as well as to her parents. One of the parents was reported to speak mostly Arabic while the other parent revealed in the interview that she often code-switched between Arabic and Swedish.

BiAra4-08’s scores on Arabic narrative comprehension were already high for MAIN1 (9 points) at Time 1 and reached ceiling at Time 2. For MAIN2, the Arabic narrative comprehension scores increased from 4 points to 7 points. For Arabic narrative production, the MAIN1 scores remained constant at 7 points while the MAIN2 scores increased by 2 points (4 points to 6 points). BiAra4-08’s Swedish scores increased on all tasks, where both MAIN1 & MAIN2 comprehension increased with 7 points. In narrative production, the child scored the highest amongst her peers at Time 2 on both narrative tasks. In the questionnaire, the parents reported that BiAra4-08 spoke ‘only or mostly Arabic’ at home. BiAra4-08’s vocabulary (CLT) scores increased from Time 1 to Time 2. Arabic CLT comprehension and production scores were higher than the Swedish ones both at 4 and age 6, but the difference between the two languages narrowed as the child got older. At Time 1, the Arabic vocabulary scores were almost double as high as the Swedish ones. Although the child was born in Sweden, her regular exposure to Swedish was reported to
have started only at age 3 which explains why the child scored so low on the Swedish vocabulary and narrative tasks at age 4.

BiAra4-14’s Arabic narrative comprehension scores increased (4 points on MAIN1 and 2 points on MAIN2) while the Arabic narrative production scores stagnated at 1 point (MAIN1) and 0 points (MAIN2). BiAra4-14’s Swedish narrative comprehension scores increased sharply by 8 points (MAIN) and 7 points (MAIN2) from Time 1 to Time 2. For Swedish narrative production, the scores increased a little (by 1 or 2 points) but remained low in general. BiAra4-14’s vocabulary (CLT) scores showed a similar pattern with a slight increase in Arabic from Time 1 to Time 2 but a sharp increase in Swedish for both comprehension and production. The child’s vocabulary scores in Swedish were higher than his Arabic vocabulary scores at both Time 1 and Time 2. BiAra4-14 was reported to have a daily even amount of exposure. At home, one parent spoke mostly Arabic, while the other spoke both Arabic and Swedish. In the parental interview, it was further revealed that the child spoke mostly Swedish at home, both with siblings and parents. At school, the child had limited exposure to Arabic. When looking at the video recordings from the testing sessions, BiAra4-14 appears to be shy in both language testing sessions and only spoke the bare minimum.

BiAra4-15’s narrative comprehension scores increased in both languages from Time 1 and Time 2. The sharpest increase was for Swedish MAIN1 (from 0 points at Time 1 to 10 points at Time 2). For narrative production, Arabic MAIN1 scores remained on the same low (2 points) after 2 years, while Arabic MAIN2 scores increased by 4 points (reaching 6 points at Time 2). For Swedish narrative production, MAIN1 scores increased by 1 point (reaching 4 points at Time 2) and MAIN2 scores increased by 4 points (from floor level to 4 points). The child’s vocabulary (CLT) scores also increased between the two testing occasions; however, the Swedish scores were always higher than the Arabic ones, both at age 4 and age 6. BiAra4-15 was reported to have an even exposure to both languages throughout her day but to speak mostly Swedish at home. Furthermore, the child often spoke English (as well as Swedish) with her sibling. In the parental interview, the parent mentioned that s/he often spoke in Arabic but occasionally switched to Swedish when the child did not understand.

BiAra4-16’s scores increased on all tasks in both languages between Time 1 and Time 2. The sharpest increase (6 points) was in Arabic narrative production of MAIN1, and the smallest increase (1 point) was in Swedish narrative production of MAIN1. In general, the child’s Swedish scores were quite low, even at age 6. As for BiAra4-16’s vocabulary (CLT) knowledge, the child scored higher at Time 2 than Time 1, however, there was no sharp increase in scores and the child still scored below 50% on both Arabic and Swedish vocabulary production. The child was reported to communicate mostly in Arabic at home; however, Swedish and English are sometimes also used. BiAra4-16’s
parents reported that the child was exposed to Arabic the majority (80%) of the time throughout the day.

BiAra4-13’s Arabic narrative comprehension scores increased by 3 points (MAIN1) and 2 points (MAIN2) from Time 1 to Time 2, however the child’s Arabic narrative production scores decreased by 4 points (MAIN1) and by 2 points (MAIN2). All of the child’s Swedish narrative comprehension and production scores increased from Time 1 to Time 2. The sharpest increase was for the narrative production of MAIN2 where the child’s scores increased from 0 points at Time 1 to 9 points at Time 2. As for the child’s vocabulary (CLT) score change, at Time 1, BiAra4-13 scored higher in Arabic vocabulary comprehension and production than in Swedish, however, at Time 2, the child’s scores had both increased and were close, nevertheless, the Arabic vocabulary scores were still a little higher. As for daily language exposure, the child was reported to be exposed to Swedish the majority (60%) of her time, but was also exposed to Arabic at school via several Arabic-speaking staff members and children. At home, the parents reported that one parent spoke only Arabic, while the other spoke mostly Arabic with occasional Swedish.

BiAra4-18’s Arabic narrative comprehension scores in MAIN1 increased sharply (by 8 points) and reached ceiling at Time 2. For MAIN2 there was also an increase of scores (by 3 points), reaching 7 points at Time 2. For Arabic narrative production, scores decreased by 2 points (from 9 points to 7 points) in MAIN1 but increased by 5 points (to reach 8 points) in MAIN2. For Swedish narrative comprehension, scores increased for both MAIN1 (by 8 points, reaching ceiling) and MAIN2 (by 4 points to reach 5 points). For Swedish narrative production, there was an increase of 7 points (reaching 8 points by age 6) for MAIN1. For MAIN2, scores remained unchanged (at 2 points). In general, the Arabic narrative macrostructure scores were higher than the Swedish ones. BiAra4-18’s vocabulary (CLT) scores at Time 1 were higher for Arabic than for Swedish by as much as 29 points, however at Time 2, both languages’ vocabulary scores had increased and the difference between the two languages narrowed. The child was reported to be exposed the majority (70%) of her day to Arabic, and attended a school where many Arabic-speaking staff members and peers were surrounding her.

BiAra4-24 scored at ceiling (10 points) on Arabic narrative comprehension of MAIN1 already at Time 1 and almost as high (9 points) at Time 2. For Arabic MAIN2, BiAra4-24’s narrative comprehension scores increased from 4 points to 8 points. As for Arabic narrative production, scores decreased for both MAIN1 (7 points at Time 1 to 4 points at Time 2) and MAIN2 (4 points at Time 1 to 3 points at Time 2). BiAra4-24’s scores did not increase for Swedish narrative comprehension in MAIN1; the child had already scored 9 points at Time 1. For Swedish MAIN2, the child’s narrative comprehension scores increased by 4 points. For Swedish narrative production, BiAra4-24’s scores increased for MAIN2 (by 4 points) but decreased by 1 point on MAIN1 (from 3 points to 2 points). BiAra4-24’s vocabulary (CLT) scores in Arabic
were already high at Time 1, and increased slightly at Time 2. The Swedish vocabulary scores increased drastically by Time 2, almost reaching the Arabic score equivalent at age 6. The child was reported to have the majority (60%) of her daily language exposure in Arabic and, at home, all family members spoke foremost in Arabic with each other.

The three latter mentioned children, BiAra4-13, BiAra4-18 and BiAra4-24, have in common that their Arabic narrative production scores decreased in both MAIN1 and MAIN2 at Time 2 (except for BiAra4-18 whose scores increased in MAIN2 at Time 2). However, upon closer inspection, all three children were among the top scoring children in narrative production of MAIN1 at Time 1 and the top scoring children at Time 2 in narrative comprehension. The reason for the decrease in the Arabic narrative production scores might be related to the testing situation rather than the children’s language abilities.\textsuperscript{210} Although no statistical analyses were done in the present study regarding the effect of the testing location/time on the children’s performance, some general observations can be made. These three children were tested during the day during Time 1, either at preschool or at Arabic MTI class. At Time 2, however, the children were seen at home, in the afternoon after school hours. In other words, the children might not have been as concentrated on the narration tasks when tested at Time 2 like when they were at Time 1. Furthermore, the testing session of BiAra4-13 occurred in the family’s living room where a younger sibling (although quiet throughout the whole testing session), was roaming curiously around the room. This might have affected the child’s narrative production performance.

As can be seen, although there is an overall increase in the narrative development skills of the children from age 4 to age 6, this increase differs between comprehension and production of narratives, between the two languages, between the two narration tasks (MAIN1 and MAIN2) and most noticeably, between one child to another. Not all scores increase very much over time, few scores increase slightly, and a few decrease (see above) while others stagnate. Patterns of score change may be related to daily language exposure and to the children’s vocabulary scores. Children with a daily language exposure leaning towards one language tend to score somewhat higher in that language on narrative macrostructure.

8.4.5 Development of narrative macrostructure: Summary

Narrative comprehension scores in the longitudinal study significantly increased at group level from Time 1 to Time 2 for both languages and for both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats). At the individual level, in Swedish, all of the scores increased, many of which quite sharply.

\textsuperscript{210} The three children in general did not show any signs of difficulty speaking in Arabic, nor did they code-switch to Swedish.
Only one child scored the same (near ceiling with 9 points) on both testing occasions. In Arabic, two children’s scores decreased by 1 point each, one child’s score stayed the same, whereas the remaining children scored higher at Time 2 than at Time 1. Comparing the children’s performance in the longitudinal sample with the cross-sectional sample, the patterns were similar showing a clear increase in mean scores from age 4 to age 6 in both studies, hence, confirming the results of the cross-sectional study.

Narrative production scores, at group level, did not increase significantly for Arabic MAIN1, and the increase in scores barely reached significance for Arabic MAIN2. For Arabic MAIN1, some of the children’s scores increased, other children’s scores stagnated and some children’s scores decreased. When inspecting the video recordings, the children whose scores decreased by Time 2 were high performers in general whose performance on the task was hampered by the testing situation rather than their language abilities. For Arabic MAIN2, the majority of the children’s scores increased by Time 2, however a few children had lower scores after two years while one child’s score stayed at floor level. For Swedish, there was a significant increase at group level for both MAIN1 and MAIN2. At the individual level, the majority of the children’s scores increased by Time 2. Only two children did not increase their Swedish production at Time 2.

Comparisons between the children’s performance at age 4 and age 6 in the longitudinal sample with those of the same aged peers in the cross-sectional sample generally confirm the language development that occurs between age 4 to 6. The scores of the children in the cross-sectional study increased more sharply in Arabic but less sharply in Swedish between ages 4 and 6 in comparison to the children in the longitudinal study.

The production of narrative components in the longitudinal study was quite similar to that of the cross-sectional study with few exceptions.

For story complexity, in Arabic MAIN1, there was no noticeable change with age in the cross-sectional and the longitudinal studies. For MAIN2, there were similarities between the two studies with the exception of full episodes (GAOs), which in the longitudinal study at age 6 did not increase. For Swedish, the children in both studies performed similarly in MAIN1 and MAIN2 in regards to increase of full episodes (GAOs) and the decrease of ‘no sequences’ by age 6.

8.4.6 Development of narrative macrostructure: Discussion

The development of narrative comprehension and narrative production of the children in the longitudinal sample resembled to a high degree the difference in scores between the 4-year-olds and the 6-year-olds in the cross-sectional sample. Similarly, the children’s production of narrative components and story complexity were near identical in the two studies, with only few exceptions. The present findings can be compared to the longitudinal results in
Öztekin (2019). Öztekin (2019) retested a sub-group of Turkish-Swedish-speaking 4-year-olds ($N = 10$) two years later (at age 6) and found, at group level, near identical patterns for vocabulary and narrative macrostructure development in the longitudinal sample as in the cross-sectional sample between the 4-year-olds and the 6-year-olds.

Concerning narrative comprehension development in the present longitudinal study, at Time 1 (age 4), the children’s MAIN1 scores were significantly higher in Arabic than in Swedish, however this difference was not significant at Time 2 (age 6). Recall that many of the 4-year-olds had more exposure to Arabic than Swedish at this young age. For narrative comprehension at Time 2, the children’s mean scores in both languages and both narration tasks were higher in the longitudinal than in the cross-sectional study. The greatest difference was in Swedish MAIN1 where the 6-year-olds in the longitudinal study had a score range of 7–10, which is much narrower than the score range of the 6-year-olds in the cross-sectional sample (2–10 points). Many of the 6-year-olds in the cross-sectional study had a late AoO of Swedish and had not been exposed to Swedish for a long time before the testing session. This might explain the wider score range in Swedish scores in the cross-sectional study.

Almost all of the children’s narrative comprehension scores increased with age in both languages, though more so in Swedish than in Arabic. Whilst this in part may be due to more developed language skills (especially in Swedish), the improvements are also likely due to general cognitive growth/better inferential understanding (recall Section 3.2). In MAIN1, several children were able to score at ceiling at Time 2 in both languages, however no child scored at ceiling in MAIN2 in either language. This result reflects the children’s performance in the cross-sectional study where a similar task effect was found (as a group, children scored significantly higher on MAIN1 than on MAIN2). The finding of a task effect for narrative comprehension is reoccurring in several studies that have used MAIN (e.g. Lindgren, 2018; Bohnacker et al., 2020; Lindgren & Bohnacker; 2020; Bohnacker & Lindgren, 2020). Öztekin (2019) too found a task effect, not only in her cross-sectional sample of 102 Turkish-Swedish-speaking 4–7-year-olds, but also in her follow-up study of 10 4-year-olds.

For narrative production, there was a development with age for Swedish scores from Time 1 to Time 2. Similar results were found in the cross-sectional study. The development of Swedish was steep as the scores were quite low at age 4 on both MAIN1 and MAIN2. For Arabic narrative production, score development was non-significant for MAIN1 and barely reaching significance for MAIN2. Similarly, in the cross-sectional study, there was no significant difference between the age groups for Arabic MAIN1, but there was a significant difference between the 4-year-olds and the 6-year-olds in MAIN2. Similar patterns of a steep increase in the majority language and less so in the home language were observed with the Turkish-Swedish-speaking children who were tested after 2 years at age 6 (Öztekin, 2019). Similarly to the present
study, Öztekin observed more noticeable gains for Swedish whereas the Turkish narrative production scores varied (increased, stagnated or decreased) between the children.

Although the children’s development in Arabic MAIN1 and MAIN2 in the longitudinal study echoes that of the cross-sectional study, it is important to keep in mind that the longitudinal sample is quite small (N = 10) and that statistical results need to be interpreted with care. When looking at the performance of the individual children in Arabic MAIN1, three high-performing children at Time 1 scored much lower at Time 2. However, when inspecting these three children’s individual video recordings of the testing session at Time 2, one notices that the performance of the children is affected by the testing situation rather than their ability to produce narratives. Had these three children’s testing conditions been more favourable (for example, had they not been tested on a weekday afternoon or had a younger sibling not been present in the room), these children might have scored differently on the tasks. Consequently, the ‘statistical significance’ of the performance of the whole group would have changed.

Regarding the children’s production of macrostructure components, the performance of the children in the longitudinal study resembled that of the children in the cross-sectional study to a large degree. The most frequently produced macrostructure components were attempts and outcomes and the least produced was internal state term as reaction. Noticeable unexpected results in children in the longitudinal study included the low production of goals on Arabic MAIN2 at Time 2. The 6-year-olds instead expressed attempts and outcomes to a large degree in MAIN2. Another noticeable unexpected result is the lack of settings in Swedish MAIN2 at Time 2. The production of setting in the cross-sectional study was higher for MAIN2 than for MAIN1. The 6-year-old children in the longitudinal study had relatively high provision rates for setting at Time 2 on all narrative tasks (e.g. BiAra4-05, Arabic: kan fi marra bsayni ‘there was once a cat’, 1 point for time; BiAra4-08, Swedish: &eh en katt som sett en fjäril på en liten träd ‘&eh a cat that saw a butterfly on a little tree’, 1 point for place), except on Swedish MAIN2 where no child mentioned time or place. Many of the children instead began their story by directly describing the mother goat rescuing her baby.

As for the production of story complexity, the patterns in both studies generally resembled each other. The most frequent complexity type was ‘no sequence’. For Swedish, there was a clear increase with age in the production of full episodes (goal-attempt-outcome, GAO), whereas in Arabic, the provision rate for GAO decreased slightly for both MAIN1 and MAIN2 from age 4 to age 6. In sum, age development in the production of macrostructure components and story complexity was clearer in Swedish than in Arabic, in parallel with the stronger age development found for Swedish versus Arabic narrative comprehension.
Finally, each child’s macrostructure performance was individually investigated in mini-case studies where various background factors and the children’s vocabulary (CLT) scores were examined in relation to the child’s narrative production and narrative comprehension scores. Low-performing children on narrative macrostructure in a certain language usually also scored low on vocabulary (CLT) and/or had a low daily exposure to that language. Other background factors that were common among low scorers included limited parental language use with the child and the child’s late AoO of Swedish. In general, language exposure seemed to have an effect on the production and comprehension of narrative macrostructure. Future studies should further investigate different factors that may boost or delay the Arabic-Swedish-speaking children’s narrative skills.
9. Cross-sectional study in Lebanon: Vocabulary

9.1 Research questions

One hundred Arabic-speaking bilingual children (4–7) participated in the Lebanese cross-sectional study. The following research questions concerning vocabulary are investigated:

- How does Arabic vocabulary develop with age?
- Is there a relationship between age of onset and Arabic comprehension and production?
- How does language use in the home affect Arabic comprehension and production?
- How do age and background factors combined affect Arabic vocabulary?

Concerning both the Lebanese and Swedish cross-sectional studies, the following research questions are investigated:

- How do the Arabic vocabulary scores in the cross-sectional samples in Lebanon and in Sweden compare?
- How do the Arabic vocabulary scores in the two samples compare when conceptual scoring is applied?

9.2 Operationalization of variables and statistical analyses

9.2.1 Operationalization of variables

Whenever possible, operationalization of the variables was kept identical to the ones used in the data from Sweden.

For Age of Onset (AoO), parents were asked to indicate the age at which their child started to receive regular input in Arabic and in a second language, i.e. English/French. For Arabic, the majority of the children started to receive regular input in Arabic before the age of 3 (0–2;11, $N=80$) while the remaining few children had regular input in Arabic starting from the age of 3.
For twelve children, no information was available for either language. Due to the small number of children who received regular input in Arabic after the age of 3 (N = 8), statistical analyses using this split were not advisable. Instead, the children were re-categorized into children who had received regular input in Arabic from birth (N = 72) vs. children who had started to receive Arabic input some time after birth (N = 16).

Concerning the AoO of the child’s second language, the majority of the children had received regular input before the age of three (0–2;11, N = 77) while the remaining children had received regular input starting from age 3 (3;0–6;0, N = 11). For purposes of comparison, the AoO for the second language was also re-categorized: children who received regular input in a second language from birth (N = 38) vs. those who started to receive input in a second language some time after birth (N = 50).

As in the data from Sweden, socio-economic status (SES) was operationalized as parental education. Parents were asked to report the highest level of education they had achieved. The International Standard Classification of Education (ISCED) 2011 (UNESCO Institute for Statistics, 2012) was used as a basis to code the parents’ answers on a 9-level scale. For each child, an averaged score was created by adding the parents’ education score and dividing the sum by two, just as for the studies in Sweden. For children who lived in a single household (N = 6) or whose parents had only filled in information for one parent (N = 4), the value of that one parent was used. Children whose both parents had not filled in an answer (missing information, N = 22) were excluded from the SES analyses.

The mean of the SES (education) level of the families (both parents combined) that participated in the Lebanese cross-sectional study (Mean = 5.58, N = 78) was higher than the mean SES in the Swedish cross-sectional study (Mean = 4.29, N = 94). Recall however that there was a higher proportion of missing data concerning parental education in Lebanon. Recall also that the children in Lebanon were foremost recruited from private schools (where school fees are relatively high).

Recall that in the cross-sectional study in Sweden, a binary split between low-SES (ISCED 0–3, non-completed primary school up to completed secondary education) and high-SES (ISCED 4–8, post-secondary non-tertiary education up to completed doctorate) was made to investigate the effect of SES.

---

211 Seven children (BiLeb4-07, BiLeb5-11, BiLeb5-16, BiLeb5-17, BiLeb6-09, BiLeb6-26, BiLeb7-01) started to hear Arabic frequently from the age of three, while one child (BiLeb6-02) started from the age of six.

212 The 16 children who started to receive frequent Arabic input some time after birth (after age 1) had the following AoO: age 1, N = 6; age 2, N = 2; age 3, N = 7; age 5, N = 1. The children were relatively evenly distributed across the age groups.

213 These 50 children had the following AoO: age 1, N = 21; age 2, N = 18; age 3, N = 8; age 4, N = 3. The children were evenly distributed across the age groups.

214 One single parent had not filled in the education level. This child was excluded from the analysis, since no SES information was available.
on the vocabulary scores. In the Lebanese cross-sectional study, such a split was not advisable since the parents had a higher mean SES than those in Sweden and the split would have resulted in very uneven group sizes in Lebanon (low-SES $N = 8$ vs. high-SES $N = 70$). Therefore, only SES as a continuous variable was statistically investigated in the Lebanese cross-sectional study.

Parents’ language use with the child was reported on a five-point scale estimating how much Arabic and a second language each parent spoke to the child. The scale ranged from ‘(almost) only Arabic’, ‘mostly Arabic, sometimes Eng/Fre’, ‘Arabic 50% – Eng/Fre 50%’, ‘mostly Eng/Fre, sometimes Arabic’ to ‘(almost) only Eng/Fre’. The children were then classified into two ways for two different types of comparison.

In the first comparison, the children were categorized into: ‘only/mostly Arabic’ ($N = 35$) if both parents spoke ‘(almost) only Arabic’ or ‘mostly Arabic, sometimes a second language’ to the child, vs. ‘other’ ($N = 54$) if parents used any other language combination at home. In the second comparison, the children were categorized into ‘only/mostly a second language’ ($N = 11$) if both parents spoke to them ‘(almost) only a second language’ or ‘mostly a second language, sometimes Arabic’ vs. ‘other’ ($N = 78$). For 11 children, there was no information provided regarding the parents’ language use with the child. These children were excluded from the analyses of parent language to child.

Regarding the child’s language use with the parents, the same five-point-scale and divisions were used as above. In the first comparison, children who spoke ‘only/mostly Arabic’ ($N = 39$) to both parents were compared with children who spoke any ‘other’ language combinations ($N = 50$). In the second comparison, children who spoke ‘only/mostly a second language’ ($N = 18$) to both parents were compared with children who spoke ‘other’ language combinations ($N = 71$).

The child’s language use with sibling(s) was estimated on a three-point-scale (‘mostly Arabic’, ‘Arabic 50% – Eng/Fre 50%’ or ‘mostly Eng/Fre’) indicating how much Arabic and a second language the child spoke with her/his sibling(s). Parents could also write down another language combination in an ‘other’ category, or specify that their child had no sibling. The children were categorized in two ways, for two different types of comparison. In the first comparison, children who spoke ‘mostly Arabic’ ($N = 21$) were compared with children who spoke ‘other’ language combinations with their sibling(s) ($N = 58$). In the second comparison, children who spoke ‘mostly the

---

215 Similarly, a three-way split: low-SES (0–3, non-completed primary school up to completed secondary, $N = 8$), vs. mid-SES (ISCED 4–5, post-secondary non-tertiary education up to completed short-cycle tertiary education, $N = 14$) vs. high-SES (ISCED 6–8, Bachelor’s level up to a completed doctorate, $N = 56$) was also not advisable due to uneven group sizes.

216 Parents could also choose an ‘other’ category where they specified whether they spoke other languages or language combinations to the child. No parent ($N = 0$) reported such other languages (e.g. Armenian).
second language (Eng/Fre)’ \((N = 17)\) were compared with those who spoke ‘other’ language combinations with their sibling(s) \((N = 62)\).

In order to investigate whether being the oldest sibling had any effect on vocabulary knowledge, parents were asked to report their child’s birth order. Children were then classified into two groups: children who were first-born (i.e. had no older sibling(s) but had younger sibling(s), \(N = 44\)) and those who had at least one older sibling \((N = 39)\). Children who did not have any siblings \((N = 8)\) or where sibling information was missing \((N = 13)\) were excluded from the analysis.

Parents were further asked whether they read with their child (joint book reading) in the two languages. Parents could choose one of four options: reading ‘almost everyday’, ‘once/twice a week’, ‘twice a month’ and ‘(almost) never’. For each language, answers were categorized into two groups: ‘reading in X often’ (‘almost every day’ or ‘once/twice a week’) or ‘reading in X rarely’ (‘twice a month’ or ‘never’) where X refers to Arabic or a second language. In this study, only the effect of parent-and-child joint reading activities in Arabic will be analysed. A little less than half of the parents \((N = 44)\) reported that they ‘often’ read Arabic stories with their child, and about one fourth of the parents \((N = 23)\) reported that they ‘rarely’ did. The proportion of missing information for this question was quite high \((N = 34)\).

Parents were also asked to estimate their child’s current daily language exposure for each language on a seven-point scale ranging from ‘95% Arabic – 5% Eng/Fre’, ‘80% Arabic – 20% Eng/Fre’, ‘60% Arabic–40% Eng/Fre’, ‘50% Arabic–50% Eng/Fre’, ‘40% Arabic–60% Eng/Fre’, ‘20% Arabic–80% Eng/Fre’ to ‘5% Arabic, 95% Eng/Fre’.\(^{217}\) The children were then categorized into ‘mostly Arabic’ (60%–90% Arabic, \(N = 32\)), ‘even exposure’ (50% Arabic–50% Eng/Fre, \(N = 30\)) or ‘mostly a second language’ (60%–90% Eng/Fre, \(N = 27\)).

Finally, the vocabulary production target words were re-scored taking into consideration the code-switched correct responses to create a conceptual score. See Section 9.3.2.3 for more details.

9.2.2 Statistical analyses

The level of significance was set at \(p < .05\) (two-tailed) for all statistical analyses. In order to compare the comprehension and production scores, paired samples t-tests were run. Age development was investigated for age groups (one-way ANOVAs and post hoc analyses with Bonferroni correction) and

\(^{217}\) Parents could also choose an ‘other’ category if their child was exposed to another language combination. No child \((N = 0)\) was assigned to such a category. Information for 11 children was missing.
linearly as age in months (simple regression analyses). The regression analyses were also visualized as regression lines on scatterplots that showed the children’s individual scores in relation to age.

What follows is a description of the statistical analyses conducted to investigate the effect of the background factors on Arabic comprehension and production.

First, background factors were analysed individually to see whether they had an effect on the Arabic vocabulary scores. Factors that had binary splits (AoO, parental language to child, child language to parents, child language to sibling(s)) were analysed using independent samples t-tests (Welch’s, equal variance not assumed), while the factor that had a three-way-split (daily language exposure) was analysed using a one-way ANOVA followed by post hoc analyses (with Bonferroni correction). The effect of the continuous variable (SES) on vocabulary scores was analysed using a Pearson correlation (two-tailed).

Second, multifactorial analyses in the form of linear regression analyses were run on Arabic comprehension and Arabic production scores separately in order to examine the combined effects of various background factors. The regression models included only independent variables that were related to language input (rather than the child’s output) and those that had proved to have a significant effect individually on the vocabulary measure in question. Age (in months) was included in the regression analyses as a control variable. For both regression models, the following is reported for the independent variables: unstandardized coefficient ($B$), standard error ($SE$), standardized coefficient ($\beta$), and the p-value ($p$). The adjusted R-squared (adjusted $R^2$), the F-statistic (F) and its p-value are also provided.

Furthermore, comparisons between the two Arabic-speaking samples (in Lebanon and in Sweden) were made using independent samples t-tests on the age groups for both vocabulary comprehension and production. Comparisons between the two samples were also made between the production scores and the conceptual production scores using independent samples t-tests. Paired samples t-tests were used to examine differences between noun production and conceptual noun production, as well as between verb production and conceptual verb production scores.

9.3 Results

This section presents the vocabulary (CLT) results for the cross-sectional study in Lebanon, first, an overview (9.3.1), followed by vocabulary scores in relation to age (9.3.2) and in relation to various background factors (9.3.3).
9.3.1 Overview of the Arabic vocabulary results

Table 9.1 displays the mean, standard deviation (SD) and ranges for Arabic comprehension and production scores for all 100 Arabic-speaking bilingual children in Lebanon.218

<table>
<thead>
<tr>
<th></th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension scores</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>50.0</td>
</tr>
<tr>
<td>SD</td>
<td>6.0</td>
</tr>
<tr>
<td>Range</td>
<td>31–60</td>
</tr>
<tr>
<td>Production scores</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.9</td>
</tr>
<tr>
<td>SD</td>
<td>10.3</td>
</tr>
<tr>
<td>Range</td>
<td>9–56</td>
</tr>
</tbody>
</table>

Note. Max comprehension score = 60 points, max production score = 60 points.

Not surprisingly, paired samples t-tests showed that the Arabic mean comprehension scores are significantly higher than the production scores, \((t(99) = 26.988, p < .001, d = 7.059, \text{large effect size})\). The comprehension scores have a narrower variation (SD = 6.0) than the production scores (SD = 10.3). The children scored more similar to each other in comprehension than in production as clearly shown in the very large score range in the Arabic production.219

Table 9.1. All ages combined (4;0–8;1, N = 100) Arabic vocabulary (CLT) scores.

9.3.2 Vocabulary scores and age

This section starts with presenting Arabic vocabulary (CLT) comprehension and production scores in relation to age. Additionally, the section contains examples of individual children who scored considerably lower than their age group (9.3.2.1). Furthermore, a comparison between Arabic vocabulary (CLT) scores in the two cross-sectional studies is made (9.3.2.2), followed by a brief investigation of the issue of code-switching (9.3.2.3). The section ends with a summary of the main results (9.3.2.4).

218 It was not possible to classify the testing locations into ‘predominate use of Arabic’ versus ‘predominate use of English and/or French’ since testing locations varied greatly in terms of language use. Hence, no statistical analyses will be reported on whether the testing location had any effect on the children’s vocabulary score.

219 Item effects were observed in the children’s responses in Arabic production. Accuracy varied for certain test items where accuracy for some items was at ceiling (for example, \(sā’a ‘a clock’\) and for others at floor level (for example, \(mīzān ‘a scale’\)). Additionally, despite continuous efforts to establish a monolingual testing session, certain test items elicited code-switching more than other items (for example, \( mushr oom in English or champignon in French instead of \(f uṭur in Arabic\)). Error analysis of the children’s non-target responses will not be reported in the present study, however, see Section 9.3.2.3 for observations regarding code-switching in the Lebanese and Swedish samples.
### 9.3.2.1 Arabic vocabulary (CLT) scores and age

Table 9.2 below shows an increase in the mean scores for both Arabic comprehension and production as the age groups become older.

**Table 9.2. Arabic vocabulary (CLT) scores for each age group in Lebanon.**

<table>
<thead>
<tr>
<th>Age Group</th>
<th>4-year-olds</th>
<th>5-year-olds</th>
<th>6-year-olds</th>
<th>7-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>18</td>
<td>29</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>44.4</td>
<td>48.3</td>
<td>50.9</td>
<td>55.4</td>
</tr>
<tr>
<td>SD</td>
<td>6.0</td>
<td>5.1</td>
<td>5.1</td>
<td>2.6</td>
</tr>
<tr>
<td>Range</td>
<td>31–52</td>
<td>36–55</td>
<td>32–59</td>
<td>51–60</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>24.3</td>
<td>27.1</td>
<td>30.8</td>
<td>41.6</td>
</tr>
<tr>
<td>SD</td>
<td>8.5</td>
<td>7.9</td>
<td>9.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Range</td>
<td>10–37</td>
<td>10–43</td>
<td>9–45</td>
<td>28–56</td>
</tr>
</tbody>
</table>

*Note.* Max comprehension score = 60 points, max production score = 60 points.

One-way ANOVA tests confirm the increase in scores across the age groups for both comprehension and production (Arabic comprehension: $F(3,96) = 18.533, p < .001, \eta^2 = .367$, large effect size; Arabic production: $F(3,96) = 17.235, p < .001, \eta^2 = .350$, large effect size).

For Arabic comprehension, post hoc analyses (with Bonferroni correction) revealed that there was a significant difference between the oldest age group and all the younger age groups (4-year-olds vs. 7-year-olds: $p < .001$; 5-year-olds vs. 7-year-olds: $p < .001$; 6-year-olds vs. 7-year-olds: $p = .007$) as well as between the 4-year-olds and the 5-year-olds ($p = .047$) and between the 4-year-olds and the 6-year-olds ($p < .001$). The difference between the 5-year-olds and the 6-year-olds was not significant ($p = .282$).

Similarly to the comprehension scores, there was a significant difference between the oldest group and all the younger age groups in Arabic production (4-year-olds vs. 7-year-olds: $p < .001$; 5-year-olds vs. 7-year-olds: $p < .001$; 6-year-olds vs. 7-year-olds: $p < .001$). No other differences between the age groups were significant.

The standard deviation (SD) for Arabic comprehension decreased gradually as the children reached 7 years of age. For Arabic production, no similar pattern was observed. The lowest comprehension score was found in the youngest age group, whereas for production, the lowest score was found in the 6-year-old group.\(^{220}\) The highest score was found in the oldest age group (7-year-olds) for both comprehension and production. Ceiling level (60 points) was reached (by one child) in Arabic comprehension but not in Arabic production.

\(^{220}\) The 4-year-old and 5-year-old groups each include the second to lowest score (10 points), whereas the lowest score was in the 6-year-old group (9 points).
In what follows, the relationship between vocabulary scores and age as a continuous variable is investigated. A simple linear regression was run to examine whether age predicted Arabic comprehension scores and the result was significant ($F(1,98) = 58.644, p < .001$) with an $R$-squared value of $R^2 = .374$, meaning that 37.4% of the variation can be explained by the child’s age. Figure 9.1 illustrates age development in the form of a regression line drawn on a scatterplot of Arabic comprehension (CLT) scores against age in months.

![Figure 9.1. Scatterplot of Arabic (CLT) comprehension scores and age in months. Dotted lines around the regression line indicate a Confidence Interval of 95%.](image)

No child scored below 50% (below 30 points) on Arabic comprehension. The majority of the children ($N = 77$) scored between 50%–90% (30–54 points), while the remaining children ($N = 23$) scored higher than 90%. Only one 7-year-old child (BiLeb7-04) scored at ceiling with 60 points on Arabic comprehension.

Next, the children’s Arabic production scores are analysed in relation to age (as a continuous variable). A simple linear regression showed a significant relationship between Arabic production scores and age ($F(1,98) = 43.477, p < .001$) with an $R$-squared value of $R^2 = .307$, meaning that 30.7% of the variation can be explained by the child’s age (in months). Figure 9.2 visualizes age development in the form of a regression line on a scatterplot of Arabic production (CLT) scores against age in months.

---

221 The children who scored higher than 90% were two 5-year-olds, eight 6-year-olds, and thirteen 7-year-olds.
While no child scored below 50% (below 30 points) in Arabic comprehension, a little less than half of the children (N = 43) scored below 30 points in Arabic production. The children who performed low on the production part belonged to all age groups. Only one child (7-year-old BiAra7-01) scored higher than 90%, and the remaining children (N = 56) scored between 50%–90%. No child scored at ceiling in Arabic production.

Next, the relationship between comprehension and production scores is investigated and visualized in Figure 9.3. A Pearson correlation (two-tailed) showed that there was a strong positive correlation between the two vocabulary measures \( r = .750, N = 100, p < .001 \). This means that children who scored high on Arabic comprehension scored also high on Arabic production.

Figure 9.3 clearly shows that no child scored in the lower two quadrants, indicating that no child scored below 50% (below 30 points) in Arabic comprehension. About half of the children were located in the top right quadrant (N = 53) scoring over 50% on both Arabic comprehension and production. The remaining children (N = 47) scored at 50% or below on Arabic production (but higher than 50% on Arabic comprehension).
In what follows, some individual children who performed considerably lower than their age group will be discussed in relation to certain background factors that might explain their low performance.

Starting with Arabic comprehension, in the 4-year-old group, four children (BiLeb4-05, BiLeb4-14, BiLeb4-09, BiLeb4-11) stick out from their peers with particularly low scores. No background questionnaire is available for BiLeb4-05, however from analysing the child’s Arabic production answers, one can surmise that the child has a lot of input in English as the majority of the responses were in the second language and not in Arabic. For both BiLeb4-14 and BiLeb4-09, the parents reported that both of them spoke ‘mostly English or French’ with the child. Furthermore, both children’s parents reported that their child had started to receive regular input in a second language from birth while the AoO of Arabic was around age 1. For BiLeb4-11, the child was reported to have even input in her languages throughout the day. In the 5-year-old group, two children (BiLeb5-21 and BiLeb5-25) scored low. Both children were reported to have the majority of their daily language exposure in a second language (60% Eng/Fre–40% Arabic). Despite the fact that both children started to receive regular input in Arabic at birth, both children’s parents reported that they spoke with their child 50%Arabic–50%Eng/Fre.222 In the 6-year-old group, the children seem to cluster together, except for one child (BiLeb6-05) who scored as low as the lowest scoring 4-year-old. No background information about this child is available; however, looking at BiLeb6-05’s Arabic production replies, the vast majority of the answers were in the

---

222 For BiLeb5-25, data for only one parent is available and the child may be growing up in a single-parent household.
child’s second language. In the 7-year-old group, no child scored particularly low or stood out from the age group.

Next, children who stood out from their age group in Arabic production are discussed. In the 4-year-olds, the children that scored the lowest in Arabic production were the same children that scored remarkably low in comprehension (BiLeb4-05, BiLeb4-14, BiLeb4-09, BiLeb4-11). See reasoning above. All four children produced the majority of responses in their second language, despite the experimenter’s repeated efforts to encourage the children to answer in Arabic.223 In the 5-year-olds, three children (BiLeb5-06, BiLeb5-20, BiLeb5-21) scored considerably lower than their age group peers. All of the three children’s parents reported that they spoke 50%Arabic–50%Eng/Fre with their child. BiLeb5-06 and BiLeb5-20 were reported to have even language exposure to both languages throughout the day while BiLeb5-21 was reported to have the majority of his daily language exposure in his second language (60% Eng/Fre–40% Arabic). Similarly to the above mentioned low-scoring 4-year-olds, these low-scoring 5-year-olds replied many times in their second language on the Arabic production task.224 In the 6-year-old group, two children (BiLeb6-20 and BiLeb6-24) stood out from their peers and scored as low as the lowest scoring 4-year-olds. In fact, BiLeb6-24 was the lowest scoring child in Arabic production in the whole Lebanese cross-sectional sample (N=100). BiLeb6-24’s parents reported that they mostly spoke with the child in the second language and that the child almost always spoke to them in the second language, rather than in Arabic. Moreover, the child was reported to have the majority of his daily language input in a second language (60% Eng/Fre–40% Arabic). As for BiLeb6-20, the parents reported that they spoke mostly Arabic with the child and that the child had even language exposure throughout the day. Both of the low scoring 6-year-olds provided most of the answers in the Arabic production task in their second language though. Finally, in the 7-year-old group, two children (BiLeb7-16 and BiLeb7-22) scored substantially lower than their peers. One of BiLeb7-16’s parents was reported to speak to the child almost only in Arabic, while the second parent spoke mostly in the second language. Also, the child was reported to be exposed the vast majority of his day to the second language (80% Eng/Fre–20% Arabic). As for BiLeb7-22, the parents reported that they spoke ‘almost only’ and ‘mostly’ Arabic, respectively, with the child. Despite the fact that the child was reported to have an AoO of Arabic at birth, BiLeb7-22 replied, like many of the low-scoring children, more often in his second language than in Arabic on the Arabic production task, despite the experimenter’s continuous efforts to create a monolingual Arabic setting.

223 When conceptual scoring was applied (meaning that, when ‘correct’ answers in the child’s second language were scored as correct), these children’s CLT production scores doubled or tripled (see Section 9.3.2.3).

224 When conceptual scoring was used for these three low-scoring 5-year-olds, the children’s production scores doubled or quadrupled.
In sum, many of the low-scoring children in Lebanon seemed to have even language exposure or the majority of their daily language exposure in their second language. However, this is not unique for the low-scoring children, since many of the high-scoring children on the Arabic vocabulary measures have a similar daily language exposure pattern. A feature that was common amongst the low-scoring children was the parents’ language use with the child. The majority of the low-scoring children did not have both of their parents speak with them predominately in Arabic.

9.3.2.2 Comparison of Arabic vocabulary scores between the Lebanese and Swedish cross-sectional studies

In this section, a comparison of the Arabic comprehension and production scores is made between the two cross-sectional studies in Lebanon and in Sweden.

Table 9.3 All ages combined (Lebanon: 4;0–8;1, N = 100; Sweden: 4;0–7;11, N = 100) Arabic vocabulary (CLT) scores in Lebanon and Sweden.

<table>
<thead>
<tr>
<th>Arabic comprehension scores</th>
<th>Lebanon</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>50.0</td>
<td>47.5</td>
</tr>
<tr>
<td>SD</td>
<td>6.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Range</td>
<td>31–60</td>
<td>25–59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arabic production scores</th>
<th>Lebanon</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>30.9</td>
<td>32.7</td>
</tr>
<tr>
<td>SD</td>
<td>10.3</td>
<td>12.3</td>
</tr>
<tr>
<td>Range</td>
<td>9–56</td>
<td>1–53</td>
</tr>
</tbody>
</table>

*Note: Max comprehension score = 60 points, max production score = 60 points.*

As can be seen in Table 9.3, the comprehension scores were (not surprisingly) higher than the production scores in both samples. Comparing the means, the children in Lebanon (as a group) had higher Arabic comprehension scores than the children in Sweden, whereas they had lower production scores than the children in Sweden. For both comprehension and production, the standard deviations (SD) and ranges were wider in Sweden than in Lebanon, suggesting that the children in Sweden were more heterogeneous.

Figure 9.4 visualizes the mean Arabic comprehension scores for each country and each age group. For a detailed comparison per age group in Lebanon and Sweden, see Table A4.1 in Appendix A4.1.

For all age groups, the children in Lebanon had higher means, less variation (SD) and a narrower range of scores than their Arabic-speaking peers in Sweden. However, independent samples t-tests revealed that there were no significant differences in the comprehension scores between the two countries except for the oldest age group (4-year-olds: \( p = .195 \); 5-year-olds: \( p = .316 \); 6-
year-olds: $p = .153$; 7-year-olds: $t(44.02) = 2.763, p = .008, d = .803$, large effect size).

Figure 9.4. Line chart of Arabic vocabulary (CLT) comprehension mean scores for all age groups in Lebanon and Sweden. Error bars indicate a Confidence Interval (CI) of 95%.

As for production, Figure 9.5 visualizes the mean Arabic production scores for each country and each age group.

Figure 9.5. Line chart of Arabic vocabulary (CLT) production mean scores for all age groups in Lebanon and Sweden. Error bars indicate a Confidence Interval (CI) of 95%.

Regarding the Arabic production scores, the age groups in Sweden had higher mean scores, but also a wider score variation (SD and range) than their peers
in Lebanon. The only exception was the oldest age group (the 7-year-olds) where the children in Lebanon had a higher mean score (Lebanon: mean = 40.7; Sweden: mean = 37.1). See Table A4.1 in Appendix A4.1 for detailed information regarding means, SD, and ranges in Arabic vocabulary for all age groups in both countries.

Independent samples t-tests revealed that there was no significant difference between the production scores in the two cross-sectional samples, except for the 5-year-old group which just reached significance (4-year-olds: \( p = .701 \); 5-year-olds: \( t(40.81) = -2.021, p = .050, d = -.568 \), moderate effect size; 6-year-olds: \( p = .273 \); 7-year-olds: \( p = .171 \)).

Next, the relation between age (in months) and Arabic vocabulary is compared for the Lebanese and Swedish samples. Table 9.4 contains the \( R^2 \) values for the relation between age and the Arabic comprehension and Arabic production scores in both cross-sectional studies.

Table 9.4. \( R^2 \) values for Arabic comprehension and production scores in relation to age in the cross-sectional studies in Lebanon and Sweden.

<table>
<thead>
<tr>
<th></th>
<th>Lebanon</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension scores x age</td>
<td>( R^2 = .374 )</td>
<td>( R^2 = .248 )</td>
</tr>
<tr>
<td>Production scores x age</td>
<td>( R^2 = .307 )</td>
<td>( R^2 = .106 )</td>
</tr>
</tbody>
</table>

As can be seen in Table 9.4, vocabulary scores (in both Arabic comprehension and production) increase more strongly with age in Lebanon than in Sweden. Also, for both countries, comprehension scores increase more strongly with age than production scores. In Lebanon, there was no great difference between the relationship of comprehension vs. production scores with age, meaning that both increased strongly. In Sweden, however, the production scores had a much weaker relationship with age in comparison to the comprehension scores.

9.3.2.3 Code-switching in Arabic production

In this section, the phenomenon of code switching from Arabic to a second language during the vocabulary (CLT) production task is investigated. Despite the experimenter’s constant emphasis on speaking merely Arabic during the Arabic testing session, some children still chose to code-switch and name the target word in their second language.

All CLT production responses were scored 1 or 0 points (see Section 4.5.1). In order to calculate the occurrences of code switching, the responses in the non-target language were re-scored. Recall that during the Arabic session, the experimenter would ask ‘and in Arabic?’ if the child provided an answer in another language. All the children’s answers were written down during the session by the experimenter. When transcribing and coding all the children’s answers, if a child had answered conceptually correct yet in a non-target language (for example, answering *grenouille* ‘frog’ in French instead of *ḍifḍaʿa*...
in Arabic), the response was coded as ‘FOR’ (meaning the child had provided a correct answer in a ‘foreign’ or second language). Such ‘FOR’ responses were then added to the Arabic correct scores creating a new variable called ‘conceptual production score’ which contained the conceptually correct answers regardless of language choice.

See Table A4.2 in Appendix A4.2 for a detailed table containing the mean, SD and range of the production scores and conceptual production scores for the children (per age group) in Lebanon and in Sweden. While there was no significant difference between the Arabic production scores (for all children combined) between the two countries \( (t(192.41) = -1.102, p = .272) \), independent samples t-tests revealed that there was a significant difference between the Lebanese and Swedish scores in regard to conceptual production \( (t(187.54) = 3.99, p < .001, d = .564, \text{moderate effect size}) \). In other words, children in the Lebanese cross-sectional study code-switched more often (i.e. provided a correct answer in their second language) than the children in Sweden. Note that all these numbers only include the code-switched ‘correct’ answers (i.e. correctly naming the target item but in another language), not all instances of code-switching.

Recall that the CLT production task was divided into noun production and verb production, each containing 30 test target words (Section 4.2.1). In order to investigate whether the children code-switched more in the noun or the verb production tasks, independent samples t-tests were run. As for noun production, children in Lebanon scored significantly higher than children in Sweden when conceptual scoring was taken into account \( (t(169.18) = 6.223, p < .011, d = .880, \text{large effect size}) \) but significantly lower when scoring in Arabic only was applied \( (t(196.26) = -3.466, p = .001, d = -.490, \text{small effect size}) \). As for verb production, there was no difference between the two countries when conceptual scoring was applied \( (t(197.95) = .755, p = .451) \). However, when only target words in Arabic were taken into account, the Lebanese children scored significantly higher in verb production \( (t(189.54) = 2.112, p = .036, d = .229, \text{small effect size}) \).

Paired samples t-tests showed that in both countries there was a significant difference between noun production vs. conceptual noun production (Lebanon: \( t(99) = -19.156, p < .001, d = -1.916, \text{large effect size} \); Sweden: \( t(99) = -4.970, p < .001, d = -.583, \text{moderate effect size} \)) as well as a significant difference between the verb production vs. conceptual verb production scores (Lebanon: \( t(99) = -4.970, p < .001, d = -.497, \text{small to moderate effect size} \); Sweden: \( t(99) = -5.470, p < .001, d = -.547, \text{moderate effect size} \)).

Comparing the effect sizes (Cohen’s \( d \) values), the difference between noun production vs. conceptual noun production in Lebanon is much larger than that in Sweden. This means that the children in Lebanon tended to code-switch nouns to their second language more than the children in Sweden.
As for the performance of the different age groups, in Lebanon, paired samples t-tests revealed that there was a significant difference (with large effect sizes and all having $p < .001$) between the noun production scores and the conceptual noun production scores at all ages.\textsuperscript{225} For verb production, there was a significant difference (with moderate effect size) between the verb production scores and the conceptual verb production scores for the 4-, 5- and 6-year-olds\textsuperscript{226} but not the 7-year-olds ($p = .076$).

Figure 9.6 visualizes the mean production scores and mean conceptual production scores for both nouns and verbs for all age groups in Lebanon.

\textbf{Figure 9.6.} Line chart of Arabic vocabulary (CLT) noun and verb production and conceptual production score development (age groups) in Lebanon. Max score = 30 points.

As for the vocabulary production scores in Sweden, there was a significant difference between production scores versus conceptual production scores for both nouns and verbs for all age groups except at age 7 where the difference ceased to exist (noun production vs. conceptual noun production: $p = .079$; verb production vs. conceptual verb production: $p = .051$).\textsuperscript{227} Figure 9.7 visualizes the mean production and mean conceptual production scores for both nouns and verbs for all age groups in Sweden.

\textsuperscript{225} Arabic noun production scores in Lebanon: Noun production vs. Conceptual noun production: 4-year-olds: $t(17) = -10.584, p < .001, d = -2.495$; 5-year-olds: $t(28) = -12.035, p < .001, d = -2.235$; 6-year-olds: $t(29) = -10.875, p < .001 d = -1.985$; 7-year-olds: $t(22) = -7.741, p < .001 d = -1.614$.


\textsuperscript{227} Arabic production scores in Sweden: Noun production vs. Conceptual noun production: 4-year-olds: $t(21) = -3.805, p = .001, d = -.811$; 5-year-olds: $t(24) = -3.113, p = .005, d = - .623$;
Figure 9.7. Line chart of Arabic vocabulary (CLT) noun and verb production score and conceptual production score development (age groups) in Sweden. Max score = 30 points.

It seems that in both samples then, older children (7-year-olds) become more aware of the ‘monolingual setting’ of the testing session and as a result code-switch to a lesser extent. However, for the 7-year-olds in Lebanon, although the amount of code-switching had decreased in comparison to other age groups, the difference between noun production scores and conceptual noun production scores was still significant.

9.3.2.4 Vocabulary scores and age: Summary

Similarly to the Swedish cross-sectional study, the children in the Lebanese cross-sectional study scored significantly higher in Arabic comprehension than in production. Both Arabic vocabulary measures correlated with each other as well as increased significantly with age. Individual children who scored differently from their age group peers were discussed in relation to certain background factors that might have contributed to their low performance. The background factor that stood out in low-performing children was parental language use with the child. In the next section, the effect that certain background factors have on vocabulary scores is investigated statistically. The children in Lebanon code-switched much more than their peers in Sweden, especially on noun production.

---

6-year-olds: \( t(28) = -3.108, p = .004, d = -.577 \). Verb production vs. Conceptual verb production: 4-year-olds: \( t(21) = -3.156, p = .005, d = -.673 \); 5-year-olds: \( t(24) = -2.898, p = .008, d = -.580 \); 6-year-olds: \( t(28) = -2.850, p = .008, d = -.529 \).
9.3.3 Vocabulary scores and background factors (and in relation to the Swedish sample)

In this section, Arabic comprehension and production scores are investigated in relation to different background factors: age of onset (9.3.3.1), socio-economic status (9.3.3.2), language use in the home (9.3.3.3) and daily language exposure (9.3.3.4). Then, two regression analyses are presented to investigate the background factors’ joint effect on vocabulary scores (9.3.3.5). The section ends with a summary of the results (9.3.3.6). Each section contains a brief comparison between how the background factor in question affects the Lebanese sample in comparison to the Swedish sample.

9.3.3.1 Age of onset (AoO)

Unlike in the Swedish sample where almost all of the children started to receive regular input in Arabic at birth, in the Lebanese sample, the AoO of Arabic was not homogeneous. While the majority of the children (N = 72) started to receive regular input in Arabic at birth, a considerable number of children started to receive regular input after age 1 (N = 16).

Independent samples t-tests revealed a significant difference between the two AoO groups’ comprehension scores (t(18.14) = 2.804, p = .012, d = .991, large effect size) and their production scores (t(20.06) = 2.964, p = .008, d = .991, large effect size). Thus, children who had received regular input in Arabic from birth had significantly higher Arabic comprehension and production scores than children who had received Arabic input later in life.

As for the AoO of a second language, the majority of the children (N = 77) had started to receive regular input in another language before the age of three, while only few children (N = 11) started to receive regular input in another language after age 3. However, independent samples t-tests showed no significant difference between the two groups in regard to Arabic comprehension (t(12.51) = .007, p = .728) and production (t(12.96) = -.745, p = .470). In the Swedish sample, however, AoO of the second language (Swedish) did have a significant effect on both comprehension and production scores (moderate and large effect size, respectively) when the effect of AoO was examined as an individual factor.

---

228 As previously mentioned, the division into AoO of before and after age 3 yielded disproportionate groups (AoO of Arabic before 3, N = 80; after 3, N = 8), which precluded meaningful statistical analysis.

229 Similarly, when an alternative division of the AoO variable was considered for the Lebanese sample, namely comparing children with AoO of a second language at birth (N = 38) vs. other (N = 50), independent samples t-tests revealed no significant difference between the two groups (Arabic comprehension (t(69.20) = -.349, p = .728), Arabic production (t(70.56) = -.1.755, p = .084).

230 But see the results of the linear multiple regression analysis (Section 6.3.3.6.1) for the combined effect of the background factors on Arabic comprehension and production scores in the Swedish cross-sectional sample.
Tables A4.3, A4.4 and A4.5 in Appendix A4.3 contain detailed descriptions of the AoO of Arabic and of a second language in relation to Arabic comprehension and production scores.

9.3.3.2 Socio-economic status (SES)

In the Lebanese sample, no binary or three-way split could be made on the SES variable due to uneven distributions amongst the split groups. When SES (parental education) was considered as a continuous variable, a Pearson correlation (two-tailed) showed that there was no correlation for Arabic comprehension ($r = -.184$, $N = 78$, $p = .106$) or Arabic production ($r = -.198$, $N = 78$, $p = .082$) with SES.

This mirrors the results obtained in the Swedish sample, where no correlations were found between SES (as a continuous variable) and any of the vocabulary measures. Unlike in the Lebanese sample, in Sweden both binary and three-way SES splits could be explored. A significant effect was only found for Arabic production scores when a binary split (high-SES vs. low-SES) was applied, but with small effect size. However, this effect disappeared in the multifactorial analysis for Arabic production when other variables were added in the regression model (Section 6.3.3.6.1).

9.3.3.3 Language use in the home

In this section, language use between the family members at home is explored. In order to investigate whether parents’ language use with the child has any effect on Arabic vocabulary scores, the children were divided into different categories.

In the first comparison, the children were categorized according to whether both parents spoke ‘mostly Arabic’ ($N = 35$) to the child or whether they spoke an ‘other’ language combination ($N = 54$).

Table 9.5. Parents’ language use with child in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th></th>
<th>Only/mostly Arabic ($N = 35$)</th>
<th>Other ($N = 54$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>51.8</td>
<td>49.0</td>
</tr>
<tr>
<td>SD</td>
<td>4.1</td>
<td>5.9</td>
</tr>
<tr>
<td>Range</td>
<td>43–60</td>
<td>34–58</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>34.7</td>
<td>27.8</td>
</tr>
<tr>
<td>SD</td>
<td>9.2</td>
<td>10.0</td>
</tr>
<tr>
<td>Range</td>
<td>13–56</td>
<td>9–49</td>
</tr>
</tbody>
</table>

Note. Max comprehension score = 60 points, max production score = 60 points.

231 The distribution of the children across the age groups was not even for ‘parental language use with child’ in the first comparison: 4-year-olds: only/mostly Arabic $N = 6$, other $N = 11$; 5-year-olds: only/mostly Arabic $N = 7$, other $N = 19$; 6-year-olds: only/mostly Arabic $N = 12$, other $N = 15$; 7-year-olds: only/mostly Arabic $N = 10$, other $N = 9$. 

330
Independent samples t-tests revealed that there was a significant difference between the two groups, both for Arabic comprehension ($t(86.48) = 2.586$, $p = .011$, $d = .521$, moderate effect size) and Arabic production ($t(77.03) = 3.368$, $p = .001$, $d = .718$, moderate to large effect size). Thus, children whose parents spoke to them ‘mostly in Arabic’ scored significantly higher in Arabic vocabulary than children whose parents spoke other language combinations.

In the second comparison, the children were categorized according to whether the parents spoke ‘mostly a second language’ ($N = 11$) to the child or an ‘other’ language combination ($N = 78$).232

Table 9.6. Parents’ language use with child in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th></th>
<th>Only/mostly a second language (N = 11)</th>
<th>Other (N = 78)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>45.1</td>
<td>50.8</td>
</tr>
<tr>
<td>SD</td>
<td>5.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Range</td>
<td>34–54</td>
<td>36–60</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>21.1</td>
<td>31.8</td>
</tr>
<tr>
<td>SD</td>
<td>8.4</td>
<td>9.8</td>
</tr>
<tr>
<td>Range</td>
<td>9–35</td>
<td>10–56</td>
</tr>
</tbody>
</table>

*Note. Max comprehension score = 60 points, max production score = 60 points.*

Independent samples t-tests showed that once again, there was a significant difference between the two groups in Arabic comprehension ($t(12.25) = -3.132$, $p = .008$, $d = -1.117$, large effect size) and Arabic production ($t(14.134) = -3.883$, $p = .002$, $d = -1.113$, large effect size). Children whose parents spoke to them ‘mostly in a second language’ scored significantly lower in Arabic vocabulary than children whose parents did not.

In both the Lebanese and the Swedish cross-sectional studies, parent language use with the child had a significant effect on the children’s Arabic vocabulary scores. In the Swedish study, no household reported that both parents spoke predominately the second language to their child. In Lebanon, children whose parents spoke to them mostly a second language (instead of Arabic) had significantly lower Arabic vocabulary scores than their peers. Note however that the 4- and 5-year-olds were over-represented in the group where parents mostly addressed the child in a second language. This uneven distribution across age may influence the results (as younger children score lower than older children).

232The distribution of the children across the age groups was not even for ‘parental language use with child’ in the second comparison: 4-year-olds: only/mostly Eng/Fre $N = 4$, other $N = 13$; 5-year-olds: only/mostly Eng/Fre $N = 3$, other $N = 23$; 6-year-olds: only/mostly Eng/Fre $N = 4$, other $N = 23$; 7-year-olds: only/mostly Eng/Fre $N = 0$, other $N = 19$. 

331
As for the child’s language use with the parents, the children were also divided into two different categories. In the first comparison, the children were categorized according to whether the child was reported to speak ‘mostly Arabic’ \((N = 39)\) with the parents or ‘other’ language combinations \((N = 50)\).

**Table 9.7.** Child’s language use with parents in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th>Arabic comprehension</th>
<th>Only/mostly Arabic ((N = 39))</th>
<th>Other ((N = 50))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>51.4</td>
<td>49.1</td>
</tr>
<tr>
<td>SD</td>
<td>4.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Range</td>
<td>37–60</td>
<td>34–58</td>
</tr>
</tbody>
</table>

**Table 9.8.** Child’s language use with parents in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th>Arabic production</th>
<th>Only/mostly a second language ((N = 18))</th>
<th>Other ((N = 71))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>34.5</td>
<td>27.4</td>
</tr>
<tr>
<td>SD</td>
<td>9.1</td>
<td>10.1</td>
</tr>
<tr>
<td>Range</td>
<td>10–51</td>
<td>9–56</td>
</tr>
</tbody>
</table>

Note. Max comprehension score = 60 points, max production score = 60 points.

Independent samples t-tests revealed that children who spoke ‘mostly Arabic’ with their parents performed significantly better in Arabic comprehension \((t(86.76) = 2.097, p = .039, d = .437, \text{small effect size})\) and Arabic production \((t(85.20) = 3.497, p = .001, d = .737, \text{moderate to large effect size})\).

In the second comparison, the children who spoke ‘mostly a second language’ \((N = 18)\) were compared to children who spoke ‘other’ language combinations \((N = 71)\) to the parents.

**Table 9.8.** Child’s language use with parents in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th>Arabic production</th>
<th>Only/mostly a second language ((N = 18))</th>
<th>Other ((N = 71))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>47.0</td>
<td>50.9</td>
</tr>
<tr>
<td>SD</td>
<td>6.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Range</td>
<td>34–57</td>
<td>37–60</td>
</tr>
</tbody>
</table>

Note. Max comprehension score = 60 points, max production score = 60 points.

Independent samples t-tests revealed that there was a significant effect on the Arabic comprehension scores \((t(22.47) = -2.425, p = .024, d = -.744, \text{moderate effect size})\) but not on the Arabic production scores \((t(24.19) = -1.725, p = .097))\).

Both comparison splits regarding children’s language use with their parents (‘only/mostly Arabic’ vs. ‘other’ and ‘only/mostly a second language’ vs. ‘other’) had an even distribution across the age groups.
In both the Lebanese and the Swedish samples, children who spoke ‘mostly Arabic’ with their parents scored significantly higher on Arabic vocabulary measures than children who did not. Children who spoke ‘mostly a second language’ with their parents were not investigated in the Swedish cross-sectional study since only few children (N = 8) were reported to do this. In the Lebanese cross-sectional study, however, the children’s comprehension scores differed significantly between these two groups. As for production scores, although there was a 4.9 point difference between the means, there was no significant statistical difference between the two groups, probably because of the large range and the relatively small size of the ‘only/mostly a second language’ group (i.e. the uneven number of children between the two groups).

In order to investigate whether a child’s language use with sibling(s) affects Arabic vocabulary scores, the children were divided into two different groups. Tables A4.6 and A4.7 in Appendix A4.4 show a detailed overview on the children’s language use with the siblings.

In the first comparison, children who spoke ‘mostly Arabic’ to the sibling(s) (N = 21) were compared to children who spoke ‘other’ language combinations (N = 58). Independent samples t-tests showed that there was no significant difference between the groups, neither for Arabic comprehension (t(35.96) = .564, p = .576) nor production (t(38.50) = .956, p = .345).

In the second comparison, children who spoke ‘mostly a second language’ to the sibling(s) (N = 17) were compared with children who spoke ‘other’ language combinations (N = 62). Similarly to the previous comparison, independent samples t-tests showed no significant difference between the two groups, neither for Arabic comprehension (t(22.35) = -1.885, p = .072) nor Arabic production (t(21.14) = -1.418, p = .171).

While the child’s language use with siblings (whether they spoke ‘mostly Arabic’ or ‘mostly a second language’) showed no significant difference for either vocabulary measure in the Lebanese cross-sectional study, children’s language use with siblings seemed to play a bigger role in the Swedish sample. Recall that in Sweden, children who spoke ‘mostly Arabic’ with their siblings had significantly higher Arabic vocabulary scores than children who did not.

---

234 The distribution of the children across the age groups was not even for ‘child language use with sibling(s)’ in the first comparison: 4-year-olds: only/mostly Arabic N = 8, other N = 5; 5-year-olds: only/mostly Arabic N = 6, other N = 20; 6-year-olds: only/mostly Arabic N = 3, other N = 20; 7-year-olds: only/mostly Arabic N = 4, other N = 13.

235 The distribution of the children across the age groups was not even for ‘child language use with sibling(s)’ in the second comparison: 4-year-olds: only/mostly Eng/Fre N = 2, other N = 11; 5-year-olds: only/mostly Eng/Fre N = 8, other N = 18; 6-year-olds: only/mostly Eng/Fre N = 3, other N = 20; 7-year-olds: only/mostly Eng/Fre N = 4, other N = 13.
The children were also classified according to whether they were first-born \((N = 44)\) or whether they had at least one older sibling \((N = 39)\), in order to investigate whether having (an) older sibling(s) affected vocabulary scores.\(^{236}\)

**Table 9.9.** Children’s Arabic vocabulary (CLT) scores in relation to birth order.

<table>
<thead>
<tr>
<th>Arabic comprehension</th>
<th>First-born (N = 44)</th>
<th>Has older sibling(s) (N = 39)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>50.7</td>
<td>49.6</td>
</tr>
<tr>
<td>SD</td>
<td>4.8</td>
<td>6.1</td>
</tr>
<tr>
<td>Range</td>
<td>37–58</td>
<td>34–60</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>31.1</td>
<td>30.9</td>
</tr>
<tr>
<td>SD</td>
<td>10.0</td>
<td>10.6</td>
</tr>
<tr>
<td>Range</td>
<td>10–56</td>
<td>9–51</td>
</tr>
</tbody>
</table>

*Note.* Max comprehension score = 60 points, max production score = 60 points.

Independent samples t-tests revealed that children who were first-born did not score significantly higher in Arabic comprehension \((t(71.95) = .926, p = .357)\) or Arabic production \((t(78.42) = .086, p = .931)\) than children who had (an) older sibling(s). These results mirror those in the Swedish sample, where there was no significant difference in Arabic vocabulary scores between children who had older siblings and those who were first born.

Parent-and-child joint reading activities are an opportunity for the child to have additional language input at home. The parents were asked to specify how often they read with their child; the answers were divided into two categories: ‘Reading in Arabic often’ \((N = 44)\) and ‘Reading in Arabic rarely’ \((N = 23)\).\(^{237}\)

**Table 9.10.** Children’s Arabic vocabulary (CLT) scores in relation to joint book reading with parents.

<table>
<thead>
<tr>
<th>Joint reading in Arabic</th>
<th>Joint reading in Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>often (N = 44)</td>
<td>rarely (N = 23)</td>
</tr>
<tr>
<td>Arabic comprehension</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>50.8</td>
</tr>
<tr>
<td>SD</td>
<td>5.0</td>
</tr>
<tr>
<td>Range</td>
<td>37–60</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32.0</td>
</tr>
<tr>
<td>SD</td>
<td>11.0</td>
</tr>
<tr>
<td>Range</td>
<td>11–56</td>
</tr>
</tbody>
</table>

*Note.* Max comprehension score = 60 points, max production score = 60 points.

\(^{236}\) The distribution of the children across the age groups was *not even* for ‘birth order’: 4-year-olds: first born \(N = 8\), has older sibling(s) \(N = 6\); 5-year-olds: first born \(N = 14\), has older sibling(s) \(N = 12\); 6-year-olds: first born \(N = 9\), has older sibling(s) \(N = 16\); 7-year-olds: first born \(N = 13\), has older sibling(s) \(N = 5\).

\(^{237}\) The distribution of the children across the age groups was *not even* for ‘joint book reading in Arabic often/rarely’: 4-year-olds: often \(N = 5\), rarely \(N = 6\); 5-year-olds: often \(N = 15\), rarely \(N = 7\); 6-year-olds: often \(N = 13\), rarely \(N = 4\); 7-year-olds: often \(N = 11\), rarely \(N = 6\).
Independent samples t-tests revealed that children whose parents read with them (joint reading) in Arabic did not score significantly higher in Arabic comprehension ($t(36.18) = -1.001, p = .323$) or Arabic production ($t(51.26) = -1.067, p = .291$) than children whose parents did not. This finding for the Lebanese sample did not resemble the effect joint reading had on Arabic vocabulary scores for the children in the Swedish cross-sectional study. In Sweden, children whose parents ‘often’ read with them scored significantly higher in Arabic vocabulary that children whose parents did not. See discussion, Section 9.4.

9.3.3.4 Daily language exposure

Current daily language exposure was split in three ways: children who were exposed to ‘mostly Arabic’ (60%–90%, $N = 32$), children who had ‘even exposure’ to both languages (50%–50%, $N = 30$) and children who had ‘mostly a second language’ (60%–90%, $N = 27$). The results are plotted in Figure 9.8 for comprehension and in Figure 9.9 for production.

![Figure 9.8. Box plots of estimated daily language exposure and Arabic comprehension scores. Error bars indicate a Confidence Interval (CI) of 95%.](image)

A one way ANOVA revealed a significant effect of daily exposure on both the Arabic comprehension scores ($F(2,86) = 3.619, p = .031, \eta^2 = .078$) and the Arabic production scores ($F(2,86) = 5.101, p = .008, \eta^2 = .106$). Post hoc analyses (with Bonferroni correction) showed that for Arabic comprehension, the only significant difference existed between the ‘mostly Arabic’ and ‘mostly Eng/Fre’ group ($p = .027$) and not between the other two groups (‘mostly Arabic’ vs. ‘even’ $p = .395$; ‘even’ vs. ‘mostly Eng/Fre’ $p = .729$). As for Arabic production, there existed a significant difference between ‘mostly Arabic’ and the remaining two groups (‘mostly Arabic’ vs. ‘even’ $p$...
In both the Lebanese and Swedish sample, daily language exposure had an effect on the children’s Arabic comprehension and production scores. While there was no difference between the three groups (‘mostly Arabic’, ‘even’, and ‘mostly Swedish’) in Arabic comprehension for the children in Sweden, there was a significant difference between the ‘mostly Arabic’ vs. ‘mostly Eng/Fre’ group in Lebanon. As for the Arabic production scores, in the Lebanese sample, the difference was between the ‘mostly Arabic’ group and the remaining two groups; whereas in Sweden, the significant difference occurred predominately between ‘mostly a second language’ and the remaining groups.

### 9.3.3.5 Multivariate analyses

This section contains two regression analyses that investigate the combined effect of background factors on Arabic comprehension and Arabic production. Age (in months) was added to the models as a control variable since it significantly affected each of the vocabulary measures. Only those background variables that had proved to be significant (individually) were included in the models. Hence, for both Arabic comprehension and Arabic production, age (in months), AoO of Arabic, parents’ language with the child (Arabic) and daily language exposure in a second language were included in the regression model.\(^{238}\) The variable ‘daily language exposure’ (‘mostly a second language’

\(^{238}\) Both the variable parent language use with the child ‘mostly Arabic’ and the variable parent language use with the child ‘mostly Eng/Fre’ were significantly associated with comprehension
vs. ‘other’) was included to investigate whether having the majority of one’s language daily exposure in a second language affected the Arabic vocabulary scores.

Variables that were related to language output (rather than input) were excluded from the analyses. These variables are child language use with parents and with siblings.

Next, multiple linear regression analyses are presented for Arabic comprehension (Table 9.11) and for Arabic production (Table 9.12). The vocabulary score is the dependent variable and the background factors are the independent variables.

Table 9.11. Linear regression model of the combined effect on Arabic vocabulary (CLT) comprehension scores in Lebanon.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>34.46</td>
<td>2.46</td>
<td>&lt; .001***</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.235</td>
<td>.03</td>
<td>.58</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>AoO of Arabic</td>
<td>3.74</td>
<td>1.12</td>
<td>.28</td>
<td>.001**</td>
</tr>
<tr>
<td>Parents’ language use with child</td>
<td>1.37</td>
<td>.92</td>
<td>.13</td>
<td>.143</td>
</tr>
<tr>
<td>Daily exposure in a second language</td>
<td>.928</td>
<td>.97</td>
<td>.08</td>
<td>.340</td>
</tr>
<tr>
<td>R(^2) (adjusted)</td>
<td>.479</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * = p < .05, ** = p < .01, *** = p < .001. F(4,80) = 20.34, p < .001.

Table 9.12. Linear regression model of the combined effect on Arabic vocabulary (CLT) production scores in Lebanon.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.28</td>
<td>5.22</td>
<td>.117</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.358</td>
<td>.07</td>
<td>.46</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>AoO of Arabic</td>
<td>6.04</td>
<td>2.38</td>
<td>.23</td>
<td>.013*</td>
</tr>
<tr>
<td>Parents’ language use with child</td>
<td>5.73</td>
<td>1.95</td>
<td>.27</td>
<td>.004**</td>
</tr>
<tr>
<td>Daily exposure in a second language</td>
<td>3.57</td>
<td>2.05</td>
<td>.16</td>
<td>.085</td>
</tr>
<tr>
<td>R(^2) (adjusted)</td>
<td>.380</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * = p < .05, ** = p < .01, *** = p < .001. F(4,80) = 13.89, p < .001.

Overall, the linear regression model for Arabic comprehension explains about 48% (adjusted \(R^2 = .479\)) of the variation, and the model for Arabic production explains 38% (adjusted \(R^2 = .380\)) of the variation as measured by the CLT between ages 4–7. Age and AoO of Arabic were significant predictors of both Arabic comprehension and production scores. Parents’ language to the child (in Arabic) was also a significant predictor of Arabic production scores. For both Arabic comprehension and production, the standardized coefficient (β and production scores, however only ‘mostly Arabic’ was put in the model to parallel to the analyses in the cross-sectional study in Sweden. When the ‘mostly Eng/Fre’ was put in the model instead, it did not enhance the model \(R^2\) (adjusted) by much (.004) for Arabic comprehension, and decreased the Arabic production model’s \(R^2\) (adjusted) substantially (.039). For neither Arabic comprehension nor production was ‘mostly Eng/Fre’ significant in the model.
beta coefficient) for age was larger than that of each of the other variables, indicating that age was a stronger predictor of Arabic vocabulary in the Lebanese cross-sectional study.

Hence, for both comprehension and production, older children and children who had an early AoO to Arabic scored higher on Arabic vocabulary (CLT) than younger children and children who had a late AoO to Arabic. Furthermore, children whose parents spoke with them only/mostly in Arabic scored higher on Arabic vocabulary production than children whose parents did not speak with them only/mostly in Arabic.

9.3.3.6 Vocabulary scores and background factors: Summary

The Arabic comprehension and production vocabulary (CLT) scores of the Lebanese cross-sectional sample were analysed in relation to various background factors (AoO of Arabic and AoO of the second language, SES, different measures of language use in the home, and current daily language exposure). The effect of each background factor in the Lebanese sample was compared with the Swedish one. Factors that proved to have a significant effect on the Arabic vocabulary scores were then included in multiple regression models to calculate the combined effect that the background factors had on the dependent variables.

The regression models showed that for Arabic comprehension, age and AoO of Arabic were significant predictors while parents’ language use with the child and daily language exposure (to the second language) was not significant. Age and AoO of Arabic could explain 48% of the variation in the sample. For Arabic production, age, AoO of Arabic and parents’ language use with the child were significant predictors while daily language exposure (to the second language) was not. In the Arabic production model, however, the $R^2$ (adjusted) value was lower than that $R^2$ in the Arabic comprehension model, meaning that the regression model for vocabulary production has a lower explanatory value.

9.4 Discussion

The results of the Lebanese cross-sectional study are discussed below both in relation to previous literature and to the Swedish cross-sectional sample.

For Arabic comprehension, no child in the Lebanese sample scored below 50% and several children scored at ceiling (or near ceiling) starting from age 5. On the other hand, Arabic production scores started quite low at age 4 (below 50%, mean score 24.3/60) and nearly doubled by age 7 (mean score 41.6/60). Arabic vocabulary comprehension and production scores increased more strongly with age in Lebanon than in Sweden. The standard deviations for all age groups on both vocabulary measures were higher in Sweden than
The effect that different background factors have on Arabic vocabulary was investigated in the Lebanese sample and compared with the effects of background factors in Sweden. Not all background factors had the same effect on vocabulary in both studies, however age was a common significant factor in both Lebanon and Sweden. In general, the older a child is, the greater the child’s lexical knowledge. Similarly, previous studies of different groups of bilingual children and monolingual children show positive correlations between age and vocabulary scores (e.g. Haman et al., 2017; Lindgren, 2018; Bohnacker et al., 2020, 2021). However, some studies have argued that background factors affect the development of expressive vocabulary of bilingual children more strongly that their receptive vocabulary (e.g. Buac el al., 2014; Thordardottir, 2011).

In Lebanon, AoO of Arabic was a significant predictor in the regression models for both vocabulary comprehension and production. Recall that AoO of Arabic, in the Lebanese study, compared children who had regular exposure to Arabic from birth with those starting from age 1. However, seven children in the Lebanese cross-sectional study were reported to have an AoO of Arabic from age 3 while one child was reported to have an AoO of Arabic from age 5. For these children, Arabic may not be considered to be their first language. None of these children scored below 50% on vocabulary comprehension, however, all of them scored well below 50% on vocabulary production with the exception of one child who scored exceptionally high on Arabic production. In comparison, in the cross-sectional study in Sweden, almost all children (98/100) had an AoO of Arabic from birth. Despite this, four children (three 4-year-olds and one 5-year-old) in Sweden scored below 50% on both Arabic vocabulary comprehension and production.

One factor that has been frequently studied in the literature (e.g. Paradis, 2011; Rowe, 2012; Bohnacker et al., 2016; Öztekin, 2019) is the effect of parents’ language use with the child. In the present dissertation, parents’ language use with the child played a significant role in both Arabic-speaking groups. In Sweden, the effect of parental language use was significant for both vocabulary comprehension and production, while in Lebanon, this effect was only significant for the production scores. A reason for this may be related to the many opportunities children in Lebanon have to hear Arabic outside the home, unlike the Arabic-speaking children in Sweden. In Sweden, Arabic is a minority language where every source of Arabic language input makes a difference. In Lebanon, however, Arabic is a majority language (or at least shares this

---

239 A 7-year-old child (BiLeb7-01) with a late AoO of Arabic (starting from age 3) scored very high on both Arabic vocabulary measures. In fact, BiLeb7-01 was the only child to score more than 90% correct on Arabic vocabulary production. The child was reported to have one parent speak with him only/mostly in Arabic, while information from the other parent was missing.
status – in some contexts, e.g. at school– with English and French) where benefits of additional language input sources at home might not show on the CLT. For example, joint book reading in Arabic was a significant predictor of vocabulary production for the children in Sweden but not in Lebanon. This does not mean that joint book reading at home does not have a positive effect on the vocabulary of children in Lebanon. It merely means that the benefits of book reading could not to be captured by the CLT results. Because Arabic input is not as easily accessible in Sweden as it is in Lebanon, every additional opportunity for a child living in Sweden to hear the home language might affect the child’s vocabulary development. For instance, although not included in the regression model, speaking Arabic with siblings proved to have a significant effect on both Arabic vocabulary measures on the children in Sweden but not on the children in Lebanon.

Effects of daily language exposure were investigated for both samples. In Sweden, having the majority of one’s daily language exposure in Swedish significantly (and negatively) affected the children’s Arabic vocabulary production (but not their vocabulary comprehension). In Lebanon, no such negative effects on Arabic vocabulary were found when the child had the majority of her/his daily language exposure in English or French.

Interestingly, in neither country did SES (parental education) have a significant effect on the children’s Arabic vocabulary. In Lebanon, the mean SES of the families was slightly higher than the SES in the families in Sweden. However, the Lebanese questionnaires had much more missing information on this question than in Sweden. Had this information been filled in, effects of SES might have emerged on the vocabulary of the children in Lebanon. On the other hand, recall that the children in Lebanon were recruited from private schools. This means that, regardless of the parents’ education level, children were receiving the same type of ‘high quality education’ (Shaaban, 2017; Bahous & Nabhani, 2008) at school. Effects of SES might be more visible if data in Lebanon were also collected from public schools or schools located in other areas than where the data collection for the present study took place.

In both cross-sectional studies in Lebanon and Sweden, regression models were run to investigate the potential combined effect of the individually significant variables.

In Lebanon, the background factors that were shown to have a significant effect on Arabic vocabulary comprehension and production were age and AoO of Arabic. In addition, parents’ language use with the child was a significant predictor of vocabulary production. The explanatory values for the regression models were $R^2$ (adjusted) = 47.9% for vocabulary comprehension and $R^2$ (adjusted) = 38.0% for vocabulary production. The fact that the explanatory value (adjusted $R^2$) was lower for vocabulary production than for vocabulary comprehension indicates that the models have a higher predictive power for receptive vocabulary than for expressive vocabulary. This means that other additional factors must lie behind the variation for Arabic production and that there
is a need to study further factors that potentially affect the vocabulary of Arabic-speaking children in Lebanon.

Furthermore, the regression models in the Lebanese sample explained less of the variance in the scores than the regression models in Sweden did. This reflects the complexity of vocabulary development in Lebanon, a country where the majority language is not the main language for (pre)school instruction.

Finally, the issue of code-switching will be discussed. In Lebanon, code-switching is the social norm (Grosjean, 1982). It is considered natural and is widespread (Bacha & Bahous, 2011; Shaaban, 2017), where children can code-switch between their languages with more or less anyone in their surroundings. This is not the case for Arabic-speaking children in Sweden. At (pre)school, the language of instruction is usually Swedish and unless there is a close-by Arabic-speaking friend or staff member, Arabic is usually not spoken at (pre)school. Outside of (pre)school, when children speak with other interlocutors at home, in the playground, or in institutions where Arabic speakers gather, Arabic-Swedish code-switching may occur more frequently.

When the Arabic-speaking children were tested in Lebanon, code-switching was very common. Many children could not keep their languages separate and had a hard time accommodating to the monolingual setting (Paradis et al., 2011) when performing the vocabulary (CLT) production task. This was especially true for the production of nouns where children in Lebanon scored significantly higher than children in Sweden when conceptual scoring was taken into account, but significantly lower when only Arabic was taken into account. In other words, children in Lebanon code-switched nouns to a higher degree than children in Sweden did. Furthermore, in the Lebanese sample, code-switched nouns occurred much more frequently than code-switched verbs. These results are in line with Khoury Aouad Saliby et al., (2017) who found in their pilot study of the Lebanese CLT that Lebanese-Arabic-speaking children \((N = 42; 5:7–7:10)\) tended to code-switch more frequently to French when naming nouns than when naming verbs.

According to the knowledge of the author, this study is the first examine the effects of several background factors on Arabic vocabulary comprehension and production for a large sample of Arabic-speaking children (4–7) in Lebanon. The lexical development of Lebanese bilingual children is an understudied topic. Further research on this topic is needed, especially since the percentages of the variance explanatory values in the present study were relatively low, even after examining several background factors that have been proven in the literature to affect the lexical development of bilingual children.

---

240 Recall that in Sweden, Arabic vocabulary comprehension \(R^2\) (adjusted) was 54.9% and Arabic vocabulary production \(R^2\) (adjusted) was 48.4%. The models had a greater explanatory value for vocabulary comprehension than for production.
10. Summary of the results

This chapter provides a summary of the results of the cross-sectional study in Sweden (10.1), the longitudinal study in Sweden (10.2) and the cross-sectional study in Lebanon (10.3).

The present dissertation had three main research questions:

- RQ1: What are the vocabulary and narrative abilities of Arabic-Swedish-speaking children (age 4–7 years) in both of the children’s languages, and which background factors measurably affect these abilities?
- RQ2: How do the vocabulary and narrative abilities of Arabic-Swedish-speaking children develop from age 4 to age 6 in the children’s two languages?
- RQ3: How do background factors affect the Arabic (L1) vocabulary development of Arabic-speaking bilingual children in Lebanon in comparison to the Arabic-Swedish-speaking children in Sweden?

10.1 The cross-sectional study in Sweden

10.1.1 Vocabulary

In Chapter 6, the lexical knowledge of 100 Arabic-Swedish-speaking children (aged 4–7 years) was investigated in both languages using the CLT as a tool to measure vocabulary comprehension and production. The main findings are summarized below.

- Age development (in months) predicted vocabulary development for comprehension and production in both the majority language, Swedish, and the home language, Arabic.
- When children of all ages were combined, there was no difference between the Arabic vocabulary total scores and the Swedish vocabulary total scores.
- As expected, vocabulary comprehension scores were higher than vocabulary production scores in both languages.
- There were strong correlations between the comprehension and production scores in each language.
- There were weak correlations between the comprehension scores of the two languages and no correlations between the production scores.
• Arabic comprehension scores varied to a lesser degree than Swedish comprehension scores, whereas Arabic production scores varied to a greater degree than the Swedish production scores. In general, for both languages, scores for vocabulary production varied more than for comprehension.
• Linear regression models were run on factors that had proved to be individually significant predictors of the vocabulary scores in each language. For Arabic, age and parents’ Arabic language use with the child predicted both comprehension and production scores. For Arabic production, joint book reading in Arabic and daily language exposure in Swedish were additional predictive factors in the model (out of which daily language exposure in Swedish was the strongest predictor). For Swedish comprehension and production, age, age of onset of Swedish and daily language exposure in Arabic were all predictive factors, the strongest of which was age.
• SES (parents’ education level) did not have any effect on the vocabulary scores in either language.

10.1.2 Narrative macrostructure
Chapter 7 included analyses of narrative macrostructure comprehension and production scores of the 100 Arabic-Swedish-speaking children (4–7) in both of their languages using MAIN as a narrative elicitation tool. The key results are summarized below for both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds /Baby Goats).

Narrative comprehension
• Group mean scores increased with age for both Arabic and Swedish. In both languages, MAIN2 scores increased more strongly with age than MAIN1 scores.
• For Arabic MAIN1, differences were found between the 4- and 6-year-olds and 4- and 7-year-olds as well as between the 5- and 7-year-olds. For MAIN2, differences were found between the 4- and 6-year-olds and 4- and 7-year-olds, and between the 5- and 6-year-olds and 5- and 7-year-olds.
• For Swedish MAIN1, a significant difference was found between the 4- and 6-year-olds and 4- and 7-year-olds. For MAIN2, differences were found between the 4-year-olds and all other age groups, and between the 7-year-olds and all the younger age groups.
• Arabic and Swedish scores correlated positively with each other in both MAIN1 and MAIN2.
• There was no significant difference between the two languages in both MAIN1 and MAIN2.
• Children scored significantly higher on MAIN1 than on MAIN2 in both languages.
• Children performed similarly on the two stories of MAIN1 (Cat and Dog) and on the two stories of MAIN2 (Baby Birds and Baby Goats).
• Response patterns for the individual comprehension questions were similar across the two languages, but differed between the two tasks (MAIN1 and MAIN2) and between the different questions. For MAIN1, questions targeting goals were answered with the highest accuracy, followed by internal states, their rationale and finally theory of mind. For MAIN2, questions probing goals and internal states were answered more accurately than questions targeting theory of mind.

Narrative production
• Arabic MAIN1 scores did not increase with age (there were no significant differences between the age groups) whereas for Arabic MAIN2, there was a significant increase between the 4- and 6-year-olds and the 4- and 7-year-olds.
• For Swedish MAIN1, there were significant differences between all age groups except between the 5- and 6-year-olds. For Swedish MAIN2, differences were found between the 7-year-olds and all the younger age groups.
• Correlations between Arabic and Swedish scores in MAIN1 and MAIN2 were either weak or nonexistent.
• There was no significant difference between the languages in MAIN1. For MAIN2, children scored significantly higher in Arabic than in Swedish.
• In both languages, there were no differences in performance between MAIN1 and MAIN2.
• The children performed similarly on the two stories of MAIN1 (Cat and Dog) and the two stories of MAIN2 (Baby Birds and Baby Goats).
• Regarding story components, in MAIN1, outcomes were the most commonly produced component type in both languages. For MAIN2, attempts were mostly produced in Arabic, whereas both attempts and outcomes were mostly produced in Swedish. For both languages, the least commonly produced components were internal states as reactions and setting. However, the production of settings increased greatly at age 7, especially in Swedish.
• Regarding story complexity, the lowest episodic complexity level ‘no sequence’ was the most frequent in both languages and in all age groups. The second most commonly produced sequence was attempt-outcome.
• The production of full episodes (goal-attempt-outcome) was rare. In Arabic MAIN2, Swedish MAIN1 and Swedish MAIN2, full episodes increased with age but were still infrequent in the oldest age group.
10.2 The longitudinal study in Sweden

Chapter 8 consisted of a follow-up study on the vocabulary (CLT) and narrative macrostructure (MAIN) of ten 4-year-old children tested two years after the first testing (i.e. at age 6). The results largely confirm the findings of the cross-sectional study, with some exceptions. The main results are listed below.

10.2.1 Vocabulary

- At group level, vocabulary comprehension and production scores increased significantly with age. The Swedish scores increased more strongly with age than the Arabic scores, but no differences were found between the two languages at age 6. There was a greater variation in the scores for Arabic than for Swedish indicating that the Swedish scores were more homogeneous at age 6.
- At the individual level, all the children’s Swedish comprehension and production scores increased with age. While all children’s Arabic comprehension scores increased, not all the children’s Arabic production scores increased with age.

10.2.2 Narrative macrostructure

- For narrative comprehension, group mean scores increased with age for both Arabic and Swedish.
- For narrative production, group mean scores in Arabic did not increase for MAIN1 and the increase barely reached significance in MAIN2. For Swedish MAIN1 and MAIN2, group mean scores did increase with age.
- At the individual level, for both narrative comprehension and production, Swedish scores increased (some quite sharply) with age, whereas Arabic scores varied (some increased while others stagnated or decreased).

10.3 The cross-sectional study in Lebanon

In Chapter 9, the lexical knowledge of 100 Arabic-speaking bilingual children (aged 4–7 years) in Lebanon was investigated using the CLT as a tool to measure Arabic vocabulary comprehension and production. The main findings were compared to the Swedish sample and are summarized below.

10.3.1 Vocabulary

- In general, the children’s Arabic vocabulary results in the Lebanese sample are in line with the Arabic results of the Arabic-Swedish-speaking sample in Sweden.
There were significant developments with age for both comprehension and production group mean scores (similar to the Swedish sample). For Arabic comprehension, there were significant differences between all age groups, except between the 5- and 6-year-olds. For Arabic production, there were significant differences between the 7-year-olds and all the younger age groups.

Arabic comprehension and Arabic production scores significantly increased with age in months (similar to the Swedish sample).

As expected, Arabic vocabulary comprehension scores were higher than vocabulary production scores (similar to the Swedish sample).

Comparing the comprehension scores of the Lebanese sample with the Swedish sample, there were no significant differences between the age groups except for the 7-year-olds where the Lebanese sample scored significantly higher.

Comparing the production scores of the Lebanese sample and the Swedish sample, there were no significant differences between the age groups except for the 5-year-olds where the Swedish sample scored significantly higher.

Arabic comprehension and production scores increased more strongly with age in Lebanon than in Sweden. In both countries, comprehension scores increased more strongly with age than production scores.

Children in the Lebanese sample code-switched many more nouns than children in the Swedish sample.

Linear regression models were run on factors that had proved to be individually significant predictors for the Arabic vocabulary scores. For comprehension, age and age of onset of Arabic were significant predictors in the models (whereas parents’ Arabic language use with the child, and daily exposure in a second language were not significant variables). For Arabic production, age, AoO of Arabic, and parents’ Arabic language use with child were significant predictors (out of which age was the strongest predictor in the model) and daily exposure in a second language was not a significant variable. (Recall that in the Swedish sample, parents’ language use with the child affected both Arabic comprehension and production whereas daily exposure to a second language affected Arabic vocabulary production.)

SES (parents’ education level) was not a predictor of the children’s Arabic vocabulary scores (similar to the Swedish sample).
11. General discussion and conclusions

This chapter contains a general discussion of the results and conclusions. First, the results of the cross-sectional study and the follow-up study conducted in Sweden are discussed (11.1), followed by the cross-sectional study in Lebanon (11.2). Then, methodological considerations and limitations (11.3) as well as suggestions for future studies (11.4) are discussed. The chapter ends with conclusions from the dissertation as a whole (11.5).

Three main research questions (RQ) were asked in the present dissertation:
- **RQ1**: What are the vocabulary and narrative abilities of Arabic-Swedish-speaking children (age 4–7 years) in both of the children’s languages, and which background factors measurably affect these abilities?
- **RQ2**: How do the vocabulary and narrative abilities of Arabic-Swedish-speaking children develop from age 4 to age 6 in the children’s two languages?
- **RQ3**: How do background factors affect the Arabic (L1) vocabulary development of Arabic-speaking bilingual children in Lebanon in comparison to the Arabic-Swedish-speaking children in Sweden?

11.1 Sweden

Arabic-Swedish-speaking bilingual children were tested for their vocabulary knowledge and narrative macrostructure. Two studies were conducted in Sweden, one cross-sectional (N = 100 children, ages 4;0–7;11) and one longitudinal (N = 10 children, 4-year-olds were tested two years later). First, a brief summary of the language tendencies of the Arabic-speaking population in Sweden is presented (11.1.1), followed by a general discussion on how vocabulary and narrative macrostructure scores develop with age (11.1.2). Then, differences between the scores in the two languages are discussed (11.1.3), followed by differences between the two language domains (11.1.4) and between the comprehension and production results (11.1.5). In the final section, the results are discussed in relation to background factors and their effect on language development (11.1.6).
11.1.1 Characteristics of the Arabic-speaking bilinguals in Sweden

Up to three generations of Arabic speakers live in Sweden as a result of mass migration waves from Arabic-speaking countries (starting from the 1980’s until recently). Some children in the Swedish cross-sectional sample were born in Arabic-speaking countries and were relative newcomers to Sweden, while others were born in Sweden to parents who were born in Sweden or have been in the country for several years. This variation is reflected in the AoO of Swedish that ranges from birth to a couple of years prior the testing occasion. In contrast, AoO of Arabic was almost exclusively from birth (recall Section 5.1.2).

In Sweden, Arabic is believed to be the second-most frequent home language (after Swedish). Arabic-speaking children in Sweden may therefore have several opportunities to speak and hear their home language outside of home, especially if they live in larger urban environments (where data for the present study was collected). Through the parental interviews in the longitudinal study as well as the experimenter’s observation and parents’ own reports, there appears to be a general desire for Arabic speakers to maintain active use of the home language (Arabic) in a society where their language is a minority language. This is also reflected in the high number of school-aged children who attend Arabic MTI in Sweden (see Section 1.2). The Arabic-speaking community may maintain the home language for several reasons, for example to participate in social activities and religious services that customarily take place in Arabic. Positive attitudes towards the Arabic language by family members and the close community might add further value to the language and lead to more enthusiasm for learning it (Pearson, 2007).

It is important to keep in mind, however, that one of the recruitment criteria for the current study was that the child could speak at least some Arabic. Some Arabic-speaking families who had signed up for the study were later excluded because their children could not complete one or several tasks in Arabic (recall Section 5.1.1, see Section 11.3.1).

Hence, despite the general will to maintain the home language Arabic, actual language use at home may not always reflect this desire. In the cross-sectional sample, the majority of the parents (90%) reported to speak mainly Arabic with each other; however, fewer parents (79%) reported that both parents spoke mainly Arabic with the child, while less than half of children (39%) were reported to speak mainly Arabic with their siblings. This suggests that there might be great differences between families and within the individual families regarding the use of Arabic versus Swedish. While some children have virtually no Swedish input at home (due to a conscious family policy or simply because other family members do not know Swedish as a result of short residency time), other children may hear more Swedish from their Arabic-speaking parents who are also proficient in Swedish and from their siblings.
In the parental interviews in the longitudinal study (at age 6), many parents reported that their children spoke more and more Swedish at home, especially with siblings. The fact that bilingual children speak the majority language more often with their siblings than with their parents as they become older and as they are more exposed to school is in line with previous literature (e.g. Bridges & Hoff, 2014; Kheirkhah & Cekaite, 2018; Rojas et al., 2016; Sorenson Duncan & Paradis, 2020b). Additionally, some parents (especially the parents of the 6- and 7-year-olds) mentioned that their children started to become more interested in learning and speaking English with siblings and friends.

The results of the cross-sectional study could have been influenced by sampling and might not be representative for the language development of all Arabic-Swedish-speaking children. For example, it cannot be generalized that the majority of Arabic-Swedish-speaking 4-year-old children in all of Sweden are more proficient in Arabic than in Swedish as was the case in the present sample (according to parental reports). The current study mainly presents results at group level or age group level. However, as shown in the scatterplots, there are many ‘outliers’ that score much higher or lower than their peers. The individual case studies showed that certain input factors were patterning together for many of the children, while at other times, the background factors could not fully explain the child’s results. In other words, the language learning journey of each child is, after all, unique.

11.1.2 Language development with age in Sweden

The present study found a general development with age in both languages for vocabulary (CLT) comprehension and production and narrative macrostructure (MAIN) comprehension and production. For all measures (except narrative production in Arabic Cat/Dog), there were significant differences between the 4-year-old group and the 7-year-old group and a positive relationship between each language measure with age in months. Still, the development with age was not quite the same for the different tasks and the two languages.

Vocabulary results increased with age for both Arabic (the minority language) and Swedish (the majority language) for comprehension and production. These results are in line with previous studies that show an increase in vocabulary scores in bilingual children with regard to age in the majority language (e.g. Cobo-Lewis et al., 2002a; Dijkstra et al., 2016; Gathercole et al., 2013; Gazuza & Hedman, 2019; Lindgren & Bohnacker, 2020). The results for Arabic are not line with studies that show small or no vocabulary gains with age in the minority language (e.g. Cobo-Lewis et al., 2002b; Gagarina et al., 2017; Gazuza & Hedman, 2019; Gathercole & Thomas, 2009; Gathercole et al., 2013; Hoff et al., 2012; Leseman, 2000; Montanari et al., 2019). In the present study, vocabulary scores at group level increased with age in the home language Arabic as well (although to a lesser
degree) than the majority language. The characteristics of the Arabic-speaking population (discussed in Section 6.4.1 and summarized in Section 11.1.1) may have contributed to the sustained development of the home language with age. The different migration patterns to Sweden (age of onset of Arabic versus that of Swedish) and the relative high amount of cumulative exposure in Arabic (inside and outside the home) may have boosted the Arabic scores.

Looking at the results of the individual children in the longitudinal study, development for Arabic vocabulary was also visible from age 4 to age 6, where all the children’s Arabic comprehension scores increased by the second testing, but only for a majority of the children in Arabic production. For Swedish vocabulary, all the children’s comprehension and production scores increased from Time 1 to Time 2. The longitudinal study (at group level) largely confirms the results obtained in the cross-sectional study, namely that there are vocabulary gains in both languages but more clearly for Swedish. For Arabic vocabulary production in the longitudinal study, the children’s scores generally increased but the score range was very wide, where some children scored near ceiling whereas others scored at floor level. For Swedish, despite the strong gains in vocabulary production for all children, no child scored at ceiling level by age 6 on the CLT. In the cross-sectional study, the scores for Swedish vocabulary production did not reach ceiling level either (score range 24–46 /60 points). In comparison, Lindgren (2018) found in her cross-sectional study that Swedish-speaking monolingual children at age 6 scored near ceiling (score range 45–56 points) on vocabulary (CLT) production. The question remains whether the Swedish score range for the Arabic-Swedish-speaking children will stay large in comparison to Swedish-speaking monolinguals when the children become older and when the Arabic-Swedish-speaking children have had more exposure to the majority language. Furthermore, will the children’s Arabic keep developing steadily as the children become older?

In general, both the cross-sectional study and the longitudinal study showed that vocabulary scores increased in Swedish more clearly with age (due to less individual variation) than the Arabic scores, and vocabulary scores increased more strongly with age for comprehension than for production. This suggests that language input factors may affect receptive and expressive vocabulary differently. See Section 11.1.3 in this regard for differences between the children’s languages and Section 11.1.6 for effects of input factors on the different vocabulary measures.

For narrative macrostructure, at group level, scores for both languages increased with age for comprehension and production. These results echo previous reports in the literature on the development of narrative comprehension and production using MAIN (e.g. Bohnacker, 2016; Bohnacker et al., 2020; Bohnacker & Lindgren, 2021; Bohnacker et al., 2022; Lindgren & Bohnacker 2022; Fiani et al., 2022; Gagarina et al., 2020; Roch et al., 2016) and other
narrative stimulus materials (e.g. Berman & Slobin, 1994; Mäkinen, 2014; Trabasso & Nickels, 1992; Trabasso et al., 1992).

For Swedish narrative production, many of the younger children scored at floor level or very low due to low proficiency in Swedish. For Arabic production in Cat/Dog, many of the younger children scored as high as the oldest age group. Development with age was thus more prominent for the majority language, Swedish, compared to the minority language, Arabic, in narrative production.

Narrative comprehension scores increased more strongly with age for Arabic than for Swedish. The fact that narrative comprehension in the minority language increased more strongly with age than in the majority language is not in line with previous studies (e.g. Bohnacker et al., 2020; Lindgren, 2018; Lindgren & Bohnacker, 2020) nor with the age development in children’s vocabulary comprehension in the present study where Swedish scores increased more strongly with age than Arabic scores. What may lie behind this surprising finding for narrative comprehension? Recalling Figures 7.2 and 7.3 (in Section 7.1.3.2) and Figures 7.4 and 7.5 (in Section 7.1.3.3), it is clear that the Swedish narrative comprehension scores are much more scattered than the Arabic scores (individual differences), more so in Baby Birds/Baby Goats (MAIN2) than in Cat/Dog (MAIN1). Additionally, more of the younger children already scored at ceiling or near ceiling in Swedish narrative comprehension than in Arabic, hence there was more ‘room’ for improvement for Arabic. Recall also that a training effect was found for Swedish narrative comprehension (Section 7.1.3.1): Children who were tested second in Swedish scored higher than children whose first testing was in Swedish. So, when the scores of all the children of the same age were combined irrespective of testing order, age development for Swedish in the entire sample becomes blurred and less noticeable.

The accuracy rates of goal comprehension increased with age but differed between the age groups, languages and tasks. The age development of each of the remaining comprehension questions was not investigated in the present study. It might also be the case that children perform differently at different ages on comprehension questions targeting internal states and theory of mind. Bohnacker and Lindgren (2020), who investigated the narrative comprehension performance of different age groups (4-, 5- and 6-year-olds) monolingual Swedish-speaking children and Swedish-English-speaking children, found a general age development for the total comprehension score (all questions combined), but also that this development differed from one question to another and from one age group to another. Similar results were found by Fiani et al. (2020) for Lebanese Arabic-French-speaking children (4–9) where accuracy rates differed between the age groups and between each of the comprehension questions. Future studies on the present data should investigate the age development of the different individual comprehension questions for the Arabic-Swedish-speaking children.
In the present study, narrative story component types and story complexity were not investigated statistically, but general observations on provision rates indicate an increase with age for both production measures and both languages. In both Arabic and Swedish and for all age groups, the children’s ability to express visible actions (attempts and outcomes) was more noticeable than components that needed to be inferred (goals and internal states), in line with previous research (e.g. Bohnacker, 2016; Bohnacker et al., 2022; Lindgren & Bohnacker, 2022; Trabasso et al., 1992). For story complexity, ‘no sequence’ was the most frequent complexity type for all age groups, even though its proportion decreased somewhat with age. This indicates that although children included story components in their narratives, they rarely combined them into a complex sequence (even at age 7). There was, however, a clear increase with age (although from a low level) in the production of complete episodes (goal-attempt-outcomes) for Swedish and Arabic Baby Birds/Baby Goats, but not Arabic Cat/Dog. In general, the older children become, the more likely they are exposed to narratives and story telling activities, especially at school in the majority language, Swedish. Older children are generally more cognitively mature and are thus better able to understand and express story contents than younger children. Hence, it comes as little surprise that the children were able to produce more complete episodes at age 7 in Swedish than in Arabic. In general, results from the longitudinal study mirror those of the cross-sectional study concerning development with age for both narrative comprehension and narrative production.

In sum, age development was not the same for the two languages nor for all tasks for the studied age range. For vocabulary, age was a stronger predictor for the majority language, Swedish, than for the minority language, Arabic. However, there were large differences between the scores of the individual children and very few scored at or near ceiling, especially for production, even at age 7. For narrative macrostructure, Swedish narrative production improved more strongly with age than Arabic, while for narrative comprehension, improvement with age was stronger for Arabic than for Swedish. For the production of story components and story complexity, there were differences between the languages where scores increased more clearly for Swedish. The verbalisation of ‘visible’ story components, i.e. those depicted in the MAIN elicitation material (attempts and outcomes), increased more noticeably with age than the verbalisation of story components that had to be inferred (goals and internal states). However, these components were well understood when explicitly probed in the comprehension questions. The production of goals (and full episodes in general) was still rudimentary at age 7 whereas the comprehension of goals was very well developed by then.

In short, although general age development could be documented for all studied language aspects for the Arabic-Swedish-speaking bilingual children, development was still far from complete, even for the oldest age group.
11.1.3 Differences between the two languages (Arabic vs. Swedish)

For the whole sample as a group, no significant differences were found between the two languages for vocabulary comprehension and production. This means that the children, taken together, perform equally well in Arabic and Swedish. At the age group level, significant differences were only found for comprehension where the 4-year-olds scored higher in Arabic. At the individual level, there were great differences between the children’s performances in their two languages. Examples of children who performed very low for their age group on the vocabulary tasks in one language (recall Sections 5.3.2.1 and 5.3.2.2) show that many of the children scored low in that language but high in the other. Lexical knowledge was thus often unevenly distributed between the languages for the individual children.

For narrative macrostructure, the children scored similarly in both languages at group level in narrative comprehension. For narrative production, an advantage of Arabic over Swedish was found in Baby Birds/Baby Goats (with a small effect size) but for Cat/Dog there were no language differences (at group level). It might be the case that the difference in the children’s languages manifests itself more strongly in one task rather than in another. Taking a closer look at the age groups, the 4-year-olds and the 6-year-olds (and not all the age groups) performed significantly higher on Baby Birds/Baby Goats in Arabic than in Swedish. One reason to why these age groups scored higher in Arabic than in Swedish could be related to characteristics of the current Arabic-Swedish-speaking sample. Recall that quite a few of the 4-year-olds and 6-year-olds had relatively short AoO of Swedish (and an AoO of Arabic from birth). For these children, telling a story in Arabic would have been much easier than telling it in a language they have not had much exposure to yet.

The finding that bilingual children (at group level) score similarly across their languages on narrative comprehension but not on narrative production is in line with previous literature (e.g. Kapalková et al., 2016; Lindgren, 2018; Öztekin, 2019). The fact that the children who scored high in one language do not necessarily score high in the other language in narrative production hints that narrative macrostructure is not merely a cognitively driven skill (against claims of the universality of macrostructure, recall Section 3.2) but rather might be affected by language factors.

Similarly to the vocabulary scores, individual children’s performance on the narrative macrostructure tasks varied and some children showed great differences between their two languages. Examples of mini case studies throughout the dissertation illustrated how different the children’s performance can be from each other, even though they share a (seemingly) similar background and similar language input.
As previously mentioned, training effects were only found for narrative comprehension where children who were tested second in Swedish scored higher in Swedish Cat/Dog and Baby Birds/Baby Goats than children whose first testing was in Swedish. Similar training effects for Swedish narrative comprehension were found by Bohnacker et al. (2020) for Turkish-Swedish-speaking children, but only for Cat/Dog. Bohnacker and colleagues explain that Swedish was the weaker language for many of the Turkish-Swedish-speaking children. In the present study, many of the children were estimated (by their parents) to have better proficiency in Arabic than in Swedish. For these children, having been tested in Arabic first served as a training opportunity for what to expect and made the children more familiar with the format of the task and the types of questions asked before performing a similar task in Swedish. Summing up, no training effect could be observed for narrative production scores in the two languages, however there was a training effect for narrative comprehension to the benefit of the weaker language. This highlights the importance of understanding a language in order to perform a task in it. Having been tested second in Swedish in the narrative comprehension task helped some children to show their full potential in Swedish, which in its turn, might have led to a balanced performance in the two languages at group level.

Finally, cognates have been proven in previous studies to have a facilitating effect for bilingual children acquiring closely related languages (e.g. Lindgren & Bohnacker, 2020; Sheng et al., 2016). Arabic and Swedish are two distant languages and the vocabulary tasks share no cognates. Thus, knowing vocabulary in one language is unlikely to facilitate vocabulary acquisition or test performance in the other language. Furthermore, there were no or only weak correlations between the two languages’ vocabulary production scores or between their narrative production scores, indicating that children who performed well in one language do not necessarily perform well in the other language.

11.1.4 Language domains (vocabulary vs. narrative macrostructure)

The current study investigated two language domains, vocabulary (measured with the CLT) and narrative macrostructure (measured with the MAIN). As shown in previous literature and confirmed in the present dissertation, vocabulary acquisition is highly dependent on language input where more input results in higher scores (e.g. Hoff et al., 2012; Pearson et al., 1997; Sorenson Duncan and Paradis, 2020b; Unsworth, 2016). Narrative macrostructure, on the other hand, has been argued to be less dependent on language input and is believed to be more cognitively driven (e.g. Pearson, 2002; Pearson & de Vil-
liers, 2006; Uccelli & Páez, 2007; Paradis et al., 2011). Hence, bilingual children are expected to perform similarly in their two languages when it comes to narrative macrostructure, but not necessarily when it comes to vocabulary (see Section 11.1.6).

The relationship between vocabulary and narrative macrostructure was not investigated statistically in the present study. However, general basic vocabulary is needed to convey meaning. If a child does not have at least some lexical knowledge, then s/he will not be able to express macrostructural elements while narrating or when answering the comprehension questions. Lindgren and Bohnacker (2020) found that the vocabulary production scores of German-Swedish-speaking children’s weaker language (German) were a significant predictor of the children’s narrative comprehension results. Furthermore, vocabulary production scores (in their respective languages) were significant predictors of Turkish-Swedish-speaking children’s narrative comprehension scores (Bohnacker et al., 2020) and narrative production scores (Bohnacker et al., 2022).

While not the same vocabulary items that are tested in the CLT are needed to answer the MAIN questions (as these probe goals, internal states and theory of mind), a low vocabulary score may be a reflection of the child’s general language proficiency. To score high on the CLT, the child needs to know those specific lexical items. To score high on the narrative macrostructure tasks, a child can express her/himself in different ways despite having a limited vocabulary (for example, by using paraphrasing or circumlocutions) to get the message across. Hence, using only one task or tool gives an incomplete picture of a child’s language knowledge since different assessment instruments reveal different aspects in a child’s language.

11.1.5 Comprehension vs. production

It is widely known that children understand much more language than what they produce themselves (e.g. Bamberg & Marchman, 1990; Bohnacker, 2016; Stein & Glenn, 1979; Trabasso et al., 1992; Trabasso & Rodkin, 1994). The current study is in line with previous research and shows that vocabulary and narrative macrostructure scores are higher for comprehension than for production for the Arabic-Swedish-speaking children.

In general, the children’s scores varied more in production than in comprehension (which had smaller score ranges). This was especially true for the vocabulary scores where production scores had, in general, larger standard deviations in both languages, indicating individual differences. However, for both vocabulary and narrative macrostructure in both languages, comprehension and production scores correlated with each other indicating there is a relationship between understanding and expressing.

The narrative comprehension scores correlated also within the tasks (MAIN1 and MAIN2) and across the languages, meaning that children who
scored high in any one of the narrative comprehension tasks also scored high in the other tasks (irrespective of language). The narrative comprehension task tests inferential understanding, hence it comes as little surprise that the narrative comprehension scores correlated with each other. Despite these correlations, there were task effects where children performed better in MAIN1 (Cat/Dog) than in MAIN2 (Baby Birds/Baby Goats). Additionally, response accuracies on certain comprehension questions show that children perform differently on seemingly identical questions in MAIN1 and MAIN2 (e.g. Bohnacker et al., 2020; Lindgren & Bohnacker, 2020; Fiani et al., 2020). For narrative production, correlations between MAIN1 and MAIN2 were also strong within the same language (but not across the languages). Also, for narrative production, there were no differences between the tasks like for narrative comprehension.

Comparing the comprehension and production of story components, the results of the present study show that the Arabic-Swedish-speaking children can answer prompted questions related to goals and internal states much better than inferring and expressing goals and internal states in spontaneous narration. Hence, children aged 4 to 7 have already acquired a macrostructural schema (narrative organisation) which they foremost use in response to answering comprehension questions rather than when telling the story unprompted (e.g. Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994). This is in line with previous research on goals and internal states using MAIN (e.g. Bohnacker, 2016; Bohnacker et al., 2020; Kapalková et al., 2016; Lindgren, 2018; Lindgren & Bohnacker, 2020; Roch et al., 2016) but also in line with the general literature on children’s narrative skills suggesting that story comprehension precedes story production (e.g. Stein & Glenn, 1979; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994). Indeed, if a child does not understand the story’s cause-and-effect relationships, the plotline as well as the intentions, thoughts and emotions of the protagonists, then s/he will not be able to produce this information when telling the story to a listener (Burris & Brown, 2014; Stein & Glenn, 1979; Trabasso & Nickels, 1992; Trabasso et al., 1992). The present study also confirms previous findings that the comprehension of internal states related to feelings/emotions is more difficult than that of goals (e.g. Bohnacker, 2016; Fiani et al., 2020; Kunnari et al., 2020; Lindgren, 2018).

Results from the longitudinal study largely confirm the results obtained in the cross-sectional study that children performed better in comprehension than in production in both vocabulary and narrative macrostructure and in both of their languages.

In sum, although comprehension and production scores correlate with each other for both vocabulary and narrative macrostructure, comprehension and production skills might develop differently with age depending on task, measure and language. Children generally perform better on comprehension tasks
rather than on production tasks. Individual differences between the children are more visible in production tasks.

11.1.6 Influence of background factors on language development

In the present dissertation, the effect of background factors was investigated statistically for the children’s vocabulary scores and impressionistically through case studies of individual children, regarding narrative macrostructure scores. For vocabulary, regression analyses showed that some language input variables have more predictive power than others, which will be discussed below. It is important, however, to keep in mind that group results may mask individual differences and that the findings of the present dissertation do not necessarily apply to ‘all’ the children.

Early versus late age of onset of the majority language Swedish (with the cut-off set at 3 years) was a significant predictor of Swedish vocabulary comprehension and production (in line with Gross et al., 2014). Thordardottir (2019) found that *cumulative amount of exposure* to a language was a better predictor than AoO for receptive vocabulary scores in the majority language. Öberg (2020), who studied 99 of the children in the present study, found that *length of exposure* to Swedish was a stronger predictor than age for Swedish vocabulary comprehension and production. In the present study, age of onset had less predictive power than age (in months). This shows that different types of statistical analyses (e.g. treating a variable as binary or continuous) yield slightly different outcomes. In the present dissertation, a binary approach to AoO was used as this is commonplace in bilingual vocabulary acquisition studies (e.g. Gagarina et al., 2017; Goldberg et al., 2008; Gross et al., 2014; Ruiz-Felter et al., 2016).

It is has repeatedly been shown that the amount of language input is a significant predictor of bilingual children’s vocabulary scores (e.g. Hoff et al., 2012; Pearson et al., 1997; Thordardottir, 2011; Unsworth, 2016). The present study investigated several ways of measuring input quantity, including estimated daily exposure to each language, how much the child hears each language from the parents and whether the child hears the home language from other sources.

The majority of the parents (79%) were reported to speak mostly Arabic at home and this was found to boost the children’s Arabic vocabulary comprehension and production and at the same time not negatively impact the children’s Swedish vocabulary (in line with Montanari et al., 2019 for the minority language, Turkish, and majority language, German). Other language input sources at home included siblings. While the children’s language use with the parents mirrored for the most part the language(s) the parents spoke with the child (i.e. mostly Arabic), with siblings there was a tendency towards speaking more the majority language (in line with Bridges & Hoff 2014;
Sorenson Duncan & Paradis, 2020b). In other words, siblings were contributing to a language shift at home from speaking almost only Arabic to introducing more Swedish. However, in the present study, no measurable advantages (or disadvantages) were found when comparing the vocabulary results of children with and without an older sibling. As previously discussed in Section 6.4.4, the reason why no such effects were observed could be related to the families’ short length of exposure to the majority language (due to relative recent relocation to Sweden). Short lengths of residency might have contributed to the upkeep of the home language among the siblings rather than inviting more majority language input at home (Paradis et al., 2020). The lack of a relationship between language input from siblings and vocabulary scores might change as length of residency increases.

As children grow older, they become more exposed to the majority language through schooling/activities, and the majority language typically develops faster than the home language (de Houwer, 2009). The minority language use is usually limited to the home environment, using an everyday register with relatively few speakers. In contrast, the school language usually offers literacy and academic terminology from more interlocutors (Montanari et al., 2019).

However, in the present Swedish sample, many of the children were reported to hear the home language outside of home as well. Those parents (N = 11) who reported that their child was mostly exposed to Arabic (80% or more) during their day also reported that their child attended a (pre)school with (one or more) Arabic-speaking school staff members. In total, 25 children were reported to hear the home language from Arabic-speaking school staff. Although there were no measurable advantages on Arabic vocabulary scores when having an Arabic-speaking staff member (recall Section 6.3.3.4), this result could be due to the fact that more of the younger children in the sample had Arabic-speaking staff members than the older children. Being exposed to Arabic at (pre)school (in the playground, for example) increases a child’s daily language exposure to Arabic in comparison to a child who attends a school where no such opportunities are found.

In general, the more a child is exposed to the home language, the greater the child’s home language proficiency becomes, which in its turn leads to more language use inviting more home language use (Pearson, 2007). The majority of the children were reported to hear Arabic from (non-Swedish-speaking) extended family. This suggests that many of the children were capable, to some extent, to hold a short conversation in Arabic only. This is in line with Genesee, Boivin and Nicoladis (1996) who found that bilingual children (N = 4, average age 2;2) can make alterations to their speech to be able to converse with monolinguals (i.e. modify their communication to suit the language characteristics of their interlocutor). Pearson (2007) suggests that there are fewer chances for bilinguals to converse in the minority language if there are no monolingual speakers of that language around. Thus, keeping
frequent contact with monolingual Arabic speakers (e.g. relatives) might play a role in the maintenance of the home language. Although no statistically significant advantages on Arabic vocabulary were detected for hearing Arabic from extended family, this could be due to the fact that there were more younger children who heard Arabic from their relatives.

Another language input factor that was investigated in the present study was the effect of parent-child shared book reading in either of the child’s languages. Results showed no significant positive effect of such book reading for Swedish vocabulary or for Arabic vocabulary comprehension, but only for Arabic vocabulary production. The reason for the lack of a visible effect on Swedish vocabulary could be related to the already high proportion of Swedish language input throughout the day via (pre)school. For Arabic, every opportunity of language input might have a measurable effect. Irrespective of whether the parents read in MSA or in vernacular Arabic, it is an opportunity for the child hear decontextualized language and engage in conversations about topics not related to the here and now. However, it was not possible in the present study to analyse the quality of the book reading activities, for example, whether the child is introduced to any low-frequency words, diverse and sophisticated vocabulary (Rowe, 2012).

Another reason as to why book reading was not a predictive variable for all vocabulary measures could be related to the actual CLT tool, which might not be sensitive enough to measure language gains resulting from being read to. Recall that CLT tests basic and concrete every-day words. Analyses of the children’s incorrect responses might give a clearer idea of which CLT words were harder than others for the children.

How much book reading a child is exposed to might vary with regard to the child’s two languages. While it is expected that children in Sweden are frequently exposed to shared book reading in Swedish at preschool, this activity might occur to a lesser degree in the homes. Some families (especially newcomers to a foreign country) might not have many books in the home, and not in Arabic either. Although public libraries are available, finding books in Arabic is not a given. Furthermore, reading with one’s child might not be a common activity that occurs in all Arabic-speaking households. Consequently, not all children might be used to telling narratives using books or picture sequences (as those used in MAIN) in the home language.

Even when book reading activities at home are relatively limited, quality language input may still be provided. Holding educational conversations or playing language-stimulating games with the child might benefit the child’s language development, at least when the child is as young as the participants of the present study. Some studies have shown that quality language input can be provided through language fostering activities including lexically rich parent-child conversations, which may even have better effects on children’s vocabulary than joint book reading (Weizman & Snow, 2001; Leseman et al., 2007).
In the present study, SES was not a significant predictor of vocabulary in Arabic, the minority language, nor in Swedish, the majority language. The fact that no effects of SES were found for minority language vocabulary echoes findings for a similar-sized group of same-aged Turkish-Swedish-speaking bilinguals in Sweden (Bohnacker et al., 2016; Bohnacker et al., 2021; Öztekin, 2019) and for bilingual children with other language combinations in international studies (Buac et al., 2014; Cobo-Lewis et al., 2002b; Leseman 2000; Prevo et al., 2014). The lack of an effect of SES on Swedish vocabulary comes as a greater surprise since it does not match studies from other countries where effects of SES are clear for the majority language (Buac et al., 2014; Calvo & Bialystok, 2014; Cobo-Lewis et al., 2002a; Leseman 2000; Meir & Armon-Lotem, 2017; Prevo et al., 2014). The lack of an effect of SES on Swedish does however match the lack of an effect of SES on Swedish in the Turkish-Swedish-speaking children (Bohnacker et al., 2016; Bohnacker et al., 2021; Öztekin, 2019).

Why were there no SES (parental education) effects in the present study? The amount of time parents spend with their child may influence vocabulary development to a higher degree than the parents’ SES. For example, if a child has highly-educated parents who spend most of their time at work without providing their child with quality language input, then the child would not benefit much from the parents’ high SES. On the other hand, if a child spends a lot of time with (at least one of the) parents, then benefits of SES might be transferable. Families who have emigrated to new countries might not be able to find a job upon arrival. However, increased quantity and quality of verbal engagement in the home language with the parents (irrespective of the parents’ SES) may positively influence the children’s vocabulary development (Hart & Risley, 1995; Hoff, 2003; Rowe, 2012). Quality language input in the home language could come from shared book reading but also from rich verbal engagement during a walk in the forest or grocery shopping.

SES can be operationalized in different ways than in the present study. Family income or parental occupation are sometimes used as proxies for SES since they tend to be associated with more opportunities for language resources through high education and extra activities that are otherwise not affordable for families with a low income. Families with high SES might have a higher income than families with low SES and thus be able to provide their children with more resources and possibilities for high quality education. The present study did not ask about income (since it might be seen as a sensitive question) but rather asked about education and occupation. However, newcomers might not be able to find a job in their profession when arriving in the host country. Furthermore, Swedish schools and preschools are affordable for all families irrespective of income. Hence, SES effects related to family income differences may be partially levelled out in a Swedish setting (at least for children as old as those in the present study).
Concerning daily language exposure, the present study confirmed previous studies that vocabulary is affected by the amount of daily exposure to the respective language. Additionally, having the majority of daily language input in Arabic might negatively affect Swedish vocabulary scores, however this factor did not have a strong predictive power. For Arabic, vocabulary production was affected negatively by having the majority of daily language input in Swedish, but not Arabic vocabulary comprehension. In fact, daily language input in Swedish had the largest predictive power among the variables for Arabic vocabulary production. This result further emphasises the importance of providing the child with enough opportunities to hear and speak the home language inside and outside the home.

The effect of daily language exposure on narrative macrostructure was investigated in a few case studies of high and low performing children. The current study invites future statistical investigations of whether there is a relationship between input patterns and narrative measures of Arabic-Swedish-speaking bilingual children (see Section 11.4). Recent studies in Sweden using MAIN found that daily language exposure did not have any significant effect on narrative production scores in the majority language, Swedish, nor the home language (Turkish: Bohnacker et al., 2022; German: Lindgren & Bohnacker, 2022).

In sum, the language input factors that influenced vocabulary were not identical across vocabulary comprehension and vocabulary production (in line with Thordardottir, 2011) nor across the two languages (in line with Montanari et al., 2019). For narrative macrostructure, observations from individual case studies of the present data hint that vocabulary and language input may influence the narrative macrostructure performance of Arabic-Swedish-speaking bilingual children but that more research is needed.

11.2 Lebanon
In Lebanon, 100 Arabic-speaking bilingual children (aged 4;0–8;1 years) having English or French as their second language participated in the cross-sectional study. The children were tested for their Arabic lexical knowledge and its relation with various background factors. Below, the characteristics of this Arabic-speaking bilingual group are summarized (11.2.1), followed by a discussion on the development of their vocabulary with age (11.2.2), a comparison between comprehension and production (11.2.3), followed by a discussion on how language input factors have affected their lexical development (11.2.4). Throughout all of these sections, a comparison is made between the two cross-sectional studies, in Lebanon and Sweden.
11.2.1 Characteristics of the Arabic-speaking bilinguals in Lebanon

The fact that bilingualism is very common in Lebanese society is reflected in the sample of the present study. The majority of the children (and their parents) were born in Lebanon. 72 children had an AoO of Arabic from birth, while 77 had an AoO of a second language before the age of three. Regarding languages spoken at home, the majority of the parents (72%) reported that they spoke mainly Arabic with each other, while only half that number reported that they mainly spoke Arabic with the child. In about one-fifth of the households, both parents were reported to speak an even amount of Arabic and a second language, while in 11% of households, both parents spoke mainly English or French (recall Section 5.3).

The majority of the children (63%) were reported to have a relatively even daily language exposure to Arabic and English/French (varying between 40%–60%). Only few children (11%) were reported to have the majority of their daily language exposure to Arabic (80%–95% Arabic) whereas 15% had the majority of their daily language exposure to English/French (80% input).

The bilingual groups in Lebanon and in Sweden differ from each other in terms of the amount of exposure to their languages. While the majority of the bilingual children in Sweden hear Arabic mainly at home, the bilingual children in Lebanon may be exposed to both their languages virtually everywhere. Hence, the two bilingual groups are exposed to their languages in different domains where use of each language differs depending on location, situation, context (Grosjean, 2015). When it comes to the formal Arabic language, MSA, unsurprisingly, children in both Sweden and Lebanon, hear MSA to a lesser degree than the colloquial Arabic variety. Noticeably, more children in the Swedish sample \((N = 64)\) were reported to hear MSA than in the Lebanese sample \((N = 46)\). While the vast majority of the children in Sweden were reported to hear MSA through TV/the internet at home, the majority of the children in Lebanon were reported to hear MSA only at school. It might be the case that families who live in a society where the home language is a minority language try their best to expose their children to the home language and the formal standard language MSA, as much as possible via media, whereas families who live in a country where MSA will eventually be taught at school do not make the same effort at home.241

In general, parental attitudes towards the Arabic language seemed to be more positive in the Swedish sample than the Lebanese sample. This does not necessarily mean that the Arabic language had less value in the eyes of the parents of the Lebanese children, but rather that the two other majority

---

241 What was not mentioned by the Lebanese parents are other probable sources where children have the opportunity to hear MSA (even if passively), for example on the TV as parents are watching the news broadcast or during religious services both of which occur in MSA (and not colloquial Lebanese Arabic).
languages in Lebanon share an equally high value. When asked in the questionnaire which language(s) the Lebanese parents thought to be the most important for their child to learn, more than half of the parents (59%) answered that all three languages (Arabic and English and French) were equally important, although their child went to a school where either English or French was the main language for instruction.

11.2.2 Arabic vocabulary development with age

In the cross-sectional study in Lebanon, Arabic vocabulary scores increased with age for both comprehension and production. This general age development is in line with the results obtained in the Swedish cross-sectional study and other studies on monolingual children (Haman et al., 2017; Lindgren, 2018) and bilingual children (e.g. Bohnacker et al., 2016; Gathercole et al., 2013; Hoff et al., 2012; Montanari et al., 2019). At age group level, the 7-year-olds scored significantly higher than the youngest age group, however only one child was able to score at ceiling (on vocabulary comprehension).

As expected, the comprehension scores were higher than the production scores. While no child scored below 50% on Arabic vocabulary comprehension, surprisingly, a little less than half of the children scored below 50% on production. Furthermore, the production scores were somewhat more scattered than the comprehension scores across the age groups. Comparing the Lebanese and the Swedish sample, the Arabic scores in Sweden were more scattered than in Lebanon.

This result is understandable, since in a way, the Lebanese cross-sectional sample was more homogeneous than the Swedish cross-sectional sample. In Sweden, the sample included both first-generation bilinguals and second-generation bilinguals with different histories of migration. As a consequence, the children’s language input in Arabic (and in Swedish) differed more in the Swedish sample. The Lebanese sample was made up of families who had a somewhat more stable life (where the vast majority of the children and their parents were born in Lebanon). The Lebanese children’s language input in Arabic was relatively constant since birth.

When comparing the Lebanese sample and the Swedish sample, the Lebanese children’s vocabulary scores increased more strongly with age than those of the Swedish children. While many of the Arabic-speaking children in Sweden hear Arabic mostly at home, children in Lebanon have more opportunities to hear Arabic outside the home. Even those few Lebanese children (15%) who were estimated to be exposed the majority (80%) of their day to their L2, are able to hear Arabic from grandparents, school staff members and friends in the play group or from the shop owner round the corner. This is not always the case for children in Sweden.

The Lebanese and the Swedish children’s vocabulary differ at the age group level. For Arabic comprehension, a significant difference between the
two samples only starts to show by age 7. Although not statistically significant, at age 7, the group mean score for Arabic production of the children in Lebanon was higher than the group mean score of the 7-year-olds in Sweden. By age 7, the Lebanese children would have become more exposed to spoken Arabic as well as started to learn to read and write in Arabic. At that age, children in Sweden would have entered their first grade, and started to receive formal (written) education in Swedish, whereas the children in Lebanon would have started to have more formal education in Arabic and are in general more exposed to Arabic throughout the day.

For Arabic production, the only significant difference between the two samples was found at age 5. One possible explanation as to why the 5-year-olds in Sweden scored significantly higher than their peers in Lebanon could simply be due to sampling. Another explanation for why the 5-year-old children in the Swedish sample scored better in vocabulary production could be related to the type of language education that children receive at that age in the respective countries. At age 5 in Lebanon, children start to learn to read and write Arabic as well as a second language. Through contacts with parents during the data collection in Lebanon, many expressed that they intentionally exposed their child to L2 prior to the start of (pre)school at age 3, so that the child will have few difficulties in understanding the teachers. (Recall that about 77% of the school education is estimated to occur in L2 rather than Arabic.) In Sweden, at age 5, Arabic-speaking children are still attending pre-school where they do not receive any formal written education in Swedish (or in any other language).

Overall, the Arabic vocabulary scores in Lebanon were not higher than the scores of the children in the Swedish sample. The answer might lie in the children’s high daily language exposure to a second language and the socially accepted code-switching behaviour amongst Lebanese children.

11.2.3 Code-switching

In the vocabulary production task, many children in the Lebanese sample code-switched to their L2, despite constant emphasis from the experimenter to speak only Arabic. At group level, the children in Lebanon code-switched much more than their Swedish peers, especially in noun production. What might be the reason for this difference? Children in Lebanon are aware that people in their surroundings understand and speak two (or more) languages and that code-switching is socially acceptable. The children tested in Sweden mostly have their Arabic input from home and are not used to speaking Arabic outside of home as much as the children in Lebanon. Although virtually all Arabic speakers in Sweden also speak Swedish (at least to some degree), the children in the Swedish sample were keener on sticking to speaking Arabic only. This could be because the children in Sweden are aware that many people do not speak or understand Arabic but Swedish only. Another probable
reason as to why the children in Sweden rarely code-switched could be that many of the families originate from Syria, Palestine and Iraq, countries where Arabic has a stronger status than in Lebanon. For example, comparing the numbers of weekly hours that first graders are taught languages at school, Syrian and Palestinian children are exposed to 8–10 hours of Arabic and 2–3 hours of English, whereas Lebanese children are exposed to 7 hours of Arabic and 7 hours of English or French (Bassiouny, 2020, pp. 262–289).

Code-switching is so common in Lebanon that it is sometimes done unconsciously. Code-switching from Arabic to English/French might also occur unknowingly for Lebanese children since they hear certain every day terms and expressions in English or French but never in Arabic (recall footnote in Section 5.3.4). For example, while collecting data in Lebanon, the experimenter insisted on asking “and in Arabic?” when the child responded in English or French. Some children replied that they did not know the Arabic word while some others replied firmly “but soleil is in Arabic!” [soleil means ‘sun’ in French]. Having gone over the CLT responses in the two samples, more “I don’t know” replies were provided in the Swedish sample than in the Lebanese sample. It could be then that the Arabic-speaking children in Sweden tend to say “I don’t know” when they don’t know the test item in Arabic rather than provide a code-switched answer. Future studies can research further the types and frequencies of incorrect answers provided by the children in both Lebanese and Swedish cross-sectional studies.

11.2.4 Influence of background factors on Arabic vocabulary development

The present study investigated the effect of language input factors on the Arabic vocabulary development of Arabic-speaking bilingual children in Lebanon. The results are also discussed in relation to the background factors investigated for the Arabic-speaking bilingual children in Sweden.

Vocabulary scores were found to be affected by AoO of Arabic. Children who were exposed to Arabic from birth performed better than children who were exposed to Arabic from age 1 or later. However, for some of the children who started to hear Arabic after age 1, Arabic might not be considered to be their first language. The vast majority of these children scored very low on vocabulary production, though none scored below 50% on vocabulary comprehension. The reason why no child in the Lebanese sample scored low on vocabulary comprehension, but a large number of children scored below 50% on vocabulary production could be related to the fact that children in Lebanon are exposed to Arabic around them, even when the communication at (pre)school is mostly in the second language and even when parents choose to speak with them mostly in the second language.
For both Arabic-speaking groups, AoO of Arabic was early. However more children had started to receive language input in Arabic from birth in the Swedish sample, whereas more children had started to receive input to a second language in Lebanon before the age of 3. Furthermore, in Sweden, more parents reported that they spoke mainly Arabic with their children than parents in the Lebanese sample.

While AoO significantly predicted both vocabulary comprehension and production, parents’ language choice with the child was a predictor merely for vocabulary production. This situation is different for the Arabic-speaking children in Sweden whose parental language use affected both Arabic vocabulary production and comprehension. In Sweden, unlike in Lebanon, the input sources for Arabic may largely be limited to parent language use at home, at least for some children.

The difference in parental language use between Lebanon and Sweden could be related to the background of the parents and the language in which they speak and were educated in. The vast majority of the Lebanese parents were born in Lebanon and had undergone their education there. This means that they are bilinguals themselves and were educated in English or French to a higher degree than in Arabic (just like their children). In contrast, the parents in Sweden were probably not raised as bilinguals in their respective Arabic-speaking countries (unless they came from Lebanon). Even if they were raised as bilinguals, they would probably not be Arabic-Swedish-speaking bilinguals like their children unless they themselves are second or third generation Arabic-speakers in Sweden.

In Lebanon, language use with siblings did not have any measurable effect on Arabic vocabulary scores, in contrast to in Sweden where Arabic language use with siblings significantly predicted Arabic comprehension and production scores. In both countries, however, having an older sibling did not benefit Arabic vocabulary. It was noticeable that only one fifth of the children in Lebanon (N = 21) were reported to speak mostly Arabic with their siblings, whereas in Sweden, the number of children who spoke mostly Arabic with their siblings was almost double (N = 39). Reasons for why fewer children in the Lebanese sample use Arabic with their siblings than in the Swedish sample could be related to differences in general language use at home between the two samples and the prevalence and general acceptance of code-switching at home and in society in Lebanon.

As for SES (parental education), no effects of SES were found on the Arabic vocabulary scores in the Lebanese sample. Irrespective of parental education, all the children attended private schools (which have high tuition fees compared to public schools) that are believed to offer higher education quality than public schools (Shaaban, 2017; Bahous & Nabhani, 2008). However, the lack of an effect of SES was also found for the Swedish sample and for the ‘twin’ BiLI-TAS study of Turkish-Swedish-speaking children in Sweden (Bohnacker et al., 2016; Bohnacker et al., 2021; Öztekin, 2019). One possible
One measure of quality language input was that of joint book reading at home. Arabic joint reading activities did not have the same measurable positive effect on children’s Arabic vocabulary in the Lebanese sample as in the Swedish cross-sectional study. The status of the Arabic language might have contributed to this result. In Sweden, Arabic is a minority language where every input source of Arabic makes a difference. In Lebanon, however, Arabic is the majority language (or at least shares this status with English and French) where the benefits of additional reading input in Arabic at home may not be as pronounced and may not show up on the CLT task(s).

When parents were asked about shared book reading activities, there was a lot of missing information in both samples. Interestingly, in Sweden, there was a preference for shared book reading in Arabic, whereas in Lebanon, parents preferred to do shared book reading in the second language (English or French) and less so in Arabic. In a related vein, Fiani et al. (2020) also found that Lebanese parents prefer telling (and reading) stories to their children in French more than in (Lebanese) Arabic or MSA. However, it seems that reading (or telling) stories at home and even at schools in Lebanon is not a very common activity in the first place, and that if it was done, it would most probably be in the child’s second language (Lassalle-Gharios, 2011).

This study has examined the effect of different language input factors on vocabulary development in Arabic-speaking bilingual children who live in two different language environments, Sweden where Arabic is a minority language, and Lebanon where Arabic is a majority language (along with English/French in certain contexts). Despite the many differences and similarities that the two language input situations have, the final regression analyses showed that two variables were strong predictors for both language settings: age and parental Arabic language use with the child. The first common variable, age, clearly shows that vocabulary comprehension and production develops as children became older (in line several studies that show a general age development for the majority language, e.g. Gagarina et al., 2017; Gauza & Hedman, 2019; Gathercole et al., 2013; Haman et al., 2017; Hoff et al., 2014). The second common variable in the two samples was parental Arabic language use with the child which had a significant effect on both comprehension and production in Sweden but only on vocabulary production in Lebanon.
These results go hand in hand with studies that show a general impact of parental language use on the child’s vocabulary development (e.g. Bohnacker et al., 2016; de Houwer, 2007; Montanari et al., 2019).

In Lebanon, society is multilingual and tolerant of code-switching, in contrast to Sweden where the majority do not speak Arabic, and hence code-switching in society may not be possible. While refugees or migrants often wish to keep the home language alive while living in a new host country, residents in Lebanon (or at least those who participated in the current study) did not appear to have the same urge, which is reflected in many of the families’ language choices at home.

The Lebanese sample was included in the present study to be able to investigate and compare the Arabic language skills of bilingual children who live in countries where Arabic has a different status and function in society. Although many of the children in Sweden were exposed to Arabic to a relatively large degree, Arabic remains the society’s minority language. Even though some children in Lebanon are exposed to a majority of their time to a second language, (Lebanese) Arabic nevertheless is the society’s majority language. Children in Lebanon have numerous opportunities to hear Arabic throughout the day, be it from extended family members, at the school playground, from the shop owner or bus driver, and from the TV. More studies are needed to study the complex interplay of background factors on the vocabulary development of Arabic-speaking bilingual children in Lebanon and in Sweden.

11.3 Methodological considerations and limitations

In this section, methodological considerations and limitations are discussed, starting with issues related to the recruitment of participants (11.3.1), followed by materials (11.3.2), questionnaire (11.3.3) and procedure (11.3.4).

11.3.1 Recruitment of participants

In the recruitment leaflet for the Swedish cross-sectional study, it was specified that the child should be able to speak both Arabic and Swedish. In other words, this study directly excluded children who did not speak ‘enough’ of the two languages. Families who had only been in Sweden for a short amount of time (less than a year) were excluded from the study since it was judged unfair to set the child up for a task (in Swedish) that s/he would most likely feel uncomfortable with due to a lack of Swedish. As a result, what is lost by this selection is the possibility to analyse the L2 language knowledge (and development) of a newcomer to Sweden. Many of the children were recruited from institutions and congregations where maintaining the Arabic language is a priority. In other words, many of the children who participated in the present
study already have parents who are keen on passing on Arabic to their children, who believe that their mother tongue has a high worth, and who choose to enrol their children in activities where the home language is spoken. Many of these children were surrounded by positive conditions that helped develop both of their languages. In comparison, in a study in Germany by Montanari et al., (2019) examining the vocabulary development of the two languages of Turkish-German-speaking children, the participants were recruited on the basis at least one Turkish-speaking parent (or caregiver). The present study narrowed down the selection of participants and does not include all types of children growing-up with Arabic and Swedish such as those who only have receptive competence in one of their languages (i.e. where the child is only able to understand the language or some of it, but not able to express her/himself in it). Hence, the present study does not contain a ‘complete’ representation of Arabic-speaking bilingual children (4–7) in Sweden. The exclusion of children with receptive knowledge in Arabic might explain the fact that there was no significant difference between the children’s languages on the vocabulary tasks at group level as well as why the Arabic vocabulary scores increased strongly with age. On the other hand, had these children been included in the study, then there would be no or very limited data from them to be analysed in the first place. One of the aims of the present study was to analyse narratives from children, hence, it may be a valid reason to exclude children who were not able to produce anything in the narrative tasks.

The recruitment leaflet also specified that the child should not have a DLD diagnosis. Had children with a DLD diagnosis been included in the present study, then the scores might have looked quite different on the scatterplots where DLD children are expected to perform lower than their same aged peers.\footnote{Öberg (2020) who compared the performance of Arabic-Swedish-speaking children with and without a DLD diagnosis ($N = 11$ and $N = 99$, respectively), found that the DLD children scored below the mean of their TD peers but well within the score range of the respective TD age groups. Öberg also noted that there was great individual variation between the children in her DLD group which she related to language exposure patterns, variation regarding strengths and weaknesses in certain language domains, and, in one case, potential misdiagnosis. The large differences in language input patterns in both the DLD and the TD group were believed to be the reason for the overlap in the vocabulary scores between the two groups (Öberg, 2020, Chapter 6).} Although the present study targeted children with typical development (according to parent report), it cannot be ruled out that none of the children in fact had language difficulties. In fact, as discussed in Öberg (2020), one of the children in the Swedish cross-sectional sample was diagnosed with DLD one year and a half after the testing session (i.e. when the child was 7 years old).

Recruitment for the cross-sectional study in Sweden started foremost via contacting 31 preschools and schools. However, because of low response rates it was decided to contact associations and congregations were Arabic speakers might gather (recall Section 4.1.1). It is thus possible that children with Arabic
input outside the home are overrepresented in the sample in Sweden. In Lebanon, the author contacted several schools, organizations and individuals before travelling to Lebanon (recall Section 4.1.3). All the children who participated in the Lebanese study attended private schools which might have been the reason why the SES of the Lebanese sample was slightly higher than that of the Swedish sample. Different outcomes might have been obtained had data been collected from families with lower SES, from public schools or from schools in other regions in Lebanon. Because of limited budget and human resources, it was not possible to test Arabic-speaking children residing in cities further away in Sweden, e.g. Malmö, or in Lebanon, e.g. Kfarmatta.

No distinction was made in the present study between first-generation refugee children and other second language learners (including second-generation refugee children). Refugee children who have moved to a new country might have had disruption of their school education in their home country as well as disruption of their education in the new country due to conflict, war, displacement and frequent relocations. Furthermore, first-generation refugee children might feel distress, having lived through negative experiences (violence, deprivation, death) which might, for example, cause stress and anxiety affecting their language learning and academic performance (see Paradis et al., 2022 for a detailed overview of the challenges of migrant children). Paradis and colleagues examined (among other language aspects), the L2 receptive vocabulary (PPVT, Dunn & Dunn, 2007) of 117 Syrian refugee children (7–14 years) who had been residing in Canada for three years at the time of testing. Paradis et al. (2022) found that the life experiences that first-generation refugees pass through may influence the children’s acquisition of the majority language (L2) in a different way than that of other bilingual children learning a second language.

11.3.2 Materials

The CLT only measures a small part of a child’s lexicon, namely the comprehension and production of certain words in two word classes (nouns and verbs) on the basis of picture selection and picture naming. The CLT does not investigate any further whether the child fully understands their meaning or whether the child can use the word in various linguistic and social contexts. Children’s vocabulary is much deeper than what has been assessed here. The children’s narrative macrostructure gives a broader idea of how children use their vocabulary (although different words were used in the narration than those tested in the CLT). Future studies could calculate the total number of words and total number of different words used in the children’s narratives to give more insight into the children’s lexical knowledge and use. The advantages of using the CLT, however, include testing a limited set of nouns and verbs for both receptive and expressive knowledge in an efficient, (relatively) fun, quick and
easy manner. Issues such as fatigue or boredom would not interfere with the testing procedure or results.

Regarding the Arabic varieties, great care was taken to ensure mutual understanding between experimenter and child in Sweden, regardless of the Arabic variety that the interlocutors spoke. Different instruction and test sheets for CLT and MAIN were prepared to match the Arabic varieties of the children (Syrian, Palestinian, Iraqi, Lebanese, i.e. the Arabic varieties that are most common in Sweden). In the vocabulary (CLT) comprehension task, children were given up to three colloquial synonyms if they showed signs of not understanding the test item. This is not considered to be a facilitative method that advantaged the children in Arabic, since the children may not have been familiar with and do not use the test item provided to them first, but rather, for example, the second one. Recall that for some test items, there were up to five different correct synonyms. It cannot be guaranteed that the test item synonyms that the experimenter gave included the relevant label for that particular child. Furthermore, some children were quick to respond without showing any sign of misunderstanding. In such cases, the experimenter cannot know whether the child would have preferred to have been given a colloquial synonym but did not say so and instead chose to provide a random answer. Knowing that different societies categorise objects and name them in different ways (Pavlenko, 2011), the work executed in preparation of the present dissertation was done to the best of the author’s ability.

11.3.3 Questionnaire

The questionnaire used in the present study was already extensive and included 36 questions. However, some background information questions were asked with greater precision than others. For example, estimated daily language exposure was measured with a 7-point scale (with the option to add an other language exposure pattern) whereas the question whether the child heard the home language from siblings, friends or extended family was asked through a multiple choice question where the parent would mark the option(s) that applied. With hindsight, more detailed follow-up questions would have been useful targeting the number of siblings, friends and extended family members that the child interacted with, as well as the extent and frequency of these interactions.

Adding more questions in relation the parents’ SES might further clarify the effect SES has on vocabulary. For example, the questionnaire could have also included a question on the language in which the parents took their highest education. In a study by Sorenson Duncan and Paradis (2020a), the mother’s level of education and the language in which that education occurred were factors found to influence maternal L2 language input to the child, which in itself was a predictor of the children’s expressive (narrative) vocabulary. In the present study, information was only available for the country in which
most part of the education was taken, not the language in which the highest degree was taken in. For example, even when education has taken place in Sweden, the language of instruction could have been in another language, most probably English. Similarly, parents who have a university degree from Lebanon or other Arabic-speaking countries, might have studied in English or French rather than in Arabic. Language skills and general knowledge gained through such studies may affect parents’ language use and literacy activities with their children.

11.3.4 Procedure

As previously mentioned, the aim of the Swedish cross-sectional and longitudinal studies was to investigate how vocabulary and narrative macrostructure develop in both of the children’s languages. That is why the experimenters strove to establish as much of a monolingual setting as possible and children were reminded to speak only one of their languages at the time of testing. Similar emphasis on a monolingual setting was applied in Lebanon where only Arabic was tested. It may have been more natural for the children in the Swedish sample to keep their two languages apart, in comparison to the children in the Lebanese sample who are accustomed to use both languages without much hinder. Some studies of bilingual children (especially in Lebanon) allow the child to answer in her/his language of choice and then calculate a conceptual score based on that reply. Had this been done in the present study, then no knowledge would have been available for each of the individual languages and how each (or in this case Arabic) develops over time.

Finally, despite the fact that the experimenter small-talks with each child prior to testing in order to ‘break the ice’ between them, it remains unclear to what extent each child is able to adapt to the testing situation. The child’s familiarity with telling stories in each language in general, or telling picture stories as presented in MAIN in particular, and the ability to adapt to the experimental setting (doing some language tasks with a stranger) might play a role in the child’s narrative performance and answering comprehension questions. As pointed out by Lindgren (2018, p. 222), factors unrelated to the child’s actual language proficiency might affect the outcome. The same holds true for the vocabulary tasks (picture selection and picture naming).

11.4 Future studies

In preparation for the current dissertation, a vast amount of data was collected, both as part of the BiLI-TAS project in Sweden and from the author’s private initiative in Lebanon. Many valuable findings have been obtained, and even more can be investigated further. Some ideas for future studies have already
been mentioned in the present chapter or throughout the dissertation while others are discussed for the first time below.

For the vocabulary data, future studies could analyse the different types of incorrect responses in the CLT (error analyses). For vocabulary production, in-depth analyses of the accuracy rates for specific test items can be made since some items showed a ceiling effect while others had a floor effect. Additionally, observations for the types of incorrect answers could be provided. This could illuminate aspects of vocabulary acquisition that are not captured by a right or wrong (1/0 points) scoring. For example, Sender and Svensson (2018), who studied the incorrect answers (non-target responses) in Swedish CLT vocabulary production by 52 Turkish-Swedish-speaking children, found that 4-year-olds, to a larger degree, used gestures and answered in Turkish (i.e. code-switched) whereas 7-year-olds more often answered with ‘I don’t know’ or provided description of the word.

A lot of information was gathered via the parental questionnaire. Further analyses can be made on the present data. Variables can be divided into other categories (for example, converted into a continuous variable) or calculated in relation to other variables (for example, analysing the effect of language input from older siblings based on the estimated preference of use, mostly Arabic or mostly Swedish, and not merely on whether the child has an older sibling or not) or joined with other variables to create an ‘input index’ (e.g. see Tuller, 2015).

For the narrative macrostructure data, future studies should include multivariate statistical analyses of the relation between narrative macrostructure (comprehension and production) with vocabulary knowledge, and background factors such as language exposure. Narrative macrostructure has been argued to be universal cross-linguistically; hence, statistically analysing the effect of vocabulary and background factors on the narrative macrostructure of the Arabic-Swedish-speaking children might give deeper insights regarding the alleged universality of narrative macrostructure across the two languages of a bilingual child. In the present dissertation, examples of individual children were discussed in relation to background factors, however, relations also need to be explored quantitatively. Additionally, similarly to the vocabulary data, deeper qualitative and quantitative analyses can be made on the children’s narrative production and comprehension answers and types of incorrect responses. For example, future studies might search for possible explanations as to why the MAIN1 scores on narrative production were so high for the younger children in Arabic, or look deeper into the children’s response types on narrative comprehension for the four MAIN stories in each language.

In the present study, data was collected from a subgroup of children two years later (at age 6) to see how their narrative production abilities have changed over the years (recall that the 7-year-olds in the present study scored far from adult-like in narrative production, see Gagarina et al. 2019). These
children could be followed up longitudinally again. Extended language versions of the CLT tasks are being prepared to test older children.

Furthermore, comparisons between the performance of different bilingual groups is possible to do within the BiLI-TAS project where data is available for other languages and language combinations. The results from the present study could be used to compare the narrative macrostructure skills of Arabic-Swedish-speaking children with a typical language development with Arabic-Swedish-speaking children with a DLD diagnosis.

The 100 children in Lebanon were also tested for their narrative macrostructure abilities in Arabic using MAIN. Due to time constraints, this data could not be included in the present dissertation. Future studies could analyse the narrative comprehension and production of these children as well as analyse the results in relation to the children’s vocabulary (CLT) scores and background factors. Additionally, comparative analyses of the narrative macrostructure performance can be done between the Arabic-speaking samples in the current dissertation.

Macrostructure is just one aspect of narrative skills. Other aspects should also be investigated, for example, character introduction and maintenance, temporality, connectives, lexicon in narratives, syntactic complexity, openings/endings, and the morale of the story. Investigating further perspectives would give a wider understanding on children’s language development.

The present study did not investigate whether there are different narrative story structures or specific linguistic differences between the two languages of the bilingual children that might have contributed to the macrostructural performance. Previous research has suggested that narrative events might be expressed differently in different languages and different cultures (see Berman & Slobin, 1994; Pavlenko, 2006; Verhoeven & Strömqvist, 2001). Comparing the performance (for macrostructure and microstructural aspects) of the children in the present study with Swedish-speaking monolinguals and Arabic-speaking monolinguals and Arabic-speaking bilinguals (data from Lebanon, for example) might be a worthy topic for future research.

Finally, studies in Lebanon interested in child bilingualism in particular (in contrast to only the Arabic language as in the present study) are encouraged to study the knowledge and development of both of the Lebanese children’s languages using the CLT and the MAIN since they were constructed with the possibility to be comparable across languages.

11.5 Conclusions

The Arabic language is spoken by hundreds of millions of people worldwide. In Sweden, Arabic has become the second most spoken home language after Swedish due to mass migration from Arabic-speaking countries. Despite these
facts, Arabic is so far an understudied language in the field of language acquisition and maintenance.

The present dissertation is the first study of its kind to investigate the vocabulary development of a large number of Arabic-speaking children in relation to numerous background factors in a Swedish and in a Lebanese context. Furthermore, this study pioneers the exploration of narrative competence of Arabic-Swedish-speaking bilinguals in Sweden.

Results from the study in Sweden suggest that families living in a country where Arabic is not the majority language should be encouraged to speak their mother tongue at home. Additionally, giving the child further opportunities to hear and speak the home language in different contexts, for example through interactive joint book reading, engaging in conversations with other Arabic speakers outside the home, or attending MTI, expands and enriches the child’s vocabulary and narrative skills in topics beyond basic conversational matters.

Results from the study in Lebanon suggest that when living in a country where Arabic is the majority language, but where the child is exposed to one or more additional languages that share an equally high status, families should also engage the child in conversations or activities where Arabic is in focus. This will make it more likely that translational equivalents are developed so that children can code-switch out of choice rather than because they do not know the words in Arabic. While some language skills will eventually develop with age, surrounding the child with positive language learning experiences and various input opportunities will boost the child’s language development.

This dissertation contributes to advancing our knowledge about the language development of children in general and of Arabic-speaking bilingual children’s lexical and narrative skills in particular. The two languages of bilingual children need not develop in similar ways across language domains. Vocabulary in particular may develop differently in the two languages of bilingual children, based on aspects of language input and the general language learning conditions that the child lives in.
ملخص عام باللغة العربية

يتفاوت عدد الناطقين ثنائي اللغة في السويد تدريجياً، وبشكل متوازي في الأونة الأخيرة، وتركز السويد على تنمية قدرة الفرد بتعلم لغتين أو أكثر. ومن بين هذه اللغات، اللغة العربية التي تعتبر في يومنا هذا اللغة الأم الأكثر تداولاً في السويد بعد اللغة السويدية. أما الجاليات المتصلة باللغة العربية بشكل أساسي من لبنان (ابتداء من الثمانينات)، ثم العراق (ابتداء من التسعينات)، وكذلك سوريا (منذ سنة 2011) بسبب الحروب، فالاتحاد بأفراد العائلة الذين آتوا منهم إلى السويد.

بالرغم من الوجود الكبير للناطقين باللغة العربية في السويد، لا يوجد العديد من الأبحاث العلمية عن تطور اللغة عند الأطفال ثنائي اللغة (العربية والسويدية). الأمر الذي أدى إلى عدة تساؤلات بحثية منها: كيف يتطور معجم الأطفال في لغتهم خلال نموهم؟ كيف يسرد الأطفال ثنائي اللغة باللغة السويدية، وكيف يتطورون هذه المهارة عبر الزمن؟

يسعى الأطفال اللغة العربية (أي لغة الأقلية) في السويد بشكل خاص في المنزل واللغة السويدية (أي لغة الأكبرية) في الحضانة/المدرسة وفي المجتمع. معنى آخر، إن سماع اللغة العربية محدود نسبيًا ومقترن على أناس معينين.

برعت الأبحاث العلمية أن تطور نحو اللغة عند الأطفال ثنائي اللغة من الممكن أن يعتمد على عوامل أساسية عدة، إن كانت داخلية (مثل عمر الطفل أو مدة ابتداء سماع لغة ثانية) أو خارجية (مثل مهارتهم تعلم اللغة في محيط، المستوى العلمي عند الأهل، فرص ممارسة اللغة بانتظام، الوعي بأهمية اللغة في المجتمع...).

اللغة العربية هي اللغة الرسمية في لبنان، لكنها تشارك في الأهمية مع اللغتين الإنجليزية والفرنسية في المجتمع، وفي المجال الأكاديمي. ينشأ الأطفال في لبنان ثنائي اللغة منذ الصغر، فمن الممكن أن يسمعوا لغتين (أو أكثر) بانتظام في المنزل والمجتمع، ومن ثم يحاط الأطفال ثنائي اللغة في لبنان بعوامل أساسية يمكن أن تؤثر على تطور لغتهم العربية خلاقياً ككيفية تأثير العوامل ذاتها على تطور نحو اللغة العربية عند الأطفال الذين يعيشون في السويد.
يُركز هذا البحث على مدى معرفة المفردات وقدرة سرد القصص.

معارضة المفردات أمر مهم جدًا لأنه يسهل عملية الفهم والتواصل بين الناس. تُظهر العديد من الأبحاث أن تطور المعجم عند الأطفال ثنائي اللغة يعتمد على العديد من العوامل الأساسية. كلما زادت فرص استخدام اللغة زاد معجم المفردات في تلك اللغة. أظهرت أبحاث خارج السويد أن تطور المعجم عند الأطفال ثنائي اللغة متعلق بعمر الطفل، والمستوى العلمي لدى أهله، ونسبة حضوره لكل من لغتيه خلال الممارسة الحياتية اليومية. أما فيما يتعلق بموضوع تطور قدرة سرد القصص، يشير العديد من الأبحاث أن هذا التطور لا يعتمد على عوامل أساسية بشكل كبير، بل على تطور "المعرفة العامة" لدى الطفل (general cognitive development). وتردده أهمية السرد لدى الأطفال عند عمر الاحتكاك بالمدرسة، حيث تُظهر الأبحاث العالمية أن مهارة السرد لها علاقة إيجابية بتسمية المعرفة الكتابية والقراءة في عمر أكبر، لذا يعتبر مهارة الأطفال في السرد وفهمه في هذا البحث.

يقوم هذا البحث بدراسة تطور اللغة عند مجموعتين من الأطفال ثنائي اللغة الذين يتكلمون اللغة العربية ولغة ثانية أخرى.

المجموعة الأولى تعيش في السويد وتعلموا اللغة العربية واللغة السويدية. وكان التركيز الأكبر في هذا البحث على تطور لغات هذه المجموعة (وهي تأتي إلى مشروع BiLI-TAS تحت إشراف البروسورة أوته بوهناكر). وركز بشكل خاص على معرفة الأطفال بالمفردات المستخدمة يوميًا، وقدرهم على سرد القصص في لغتيه. وقد أدى الأطفال في المجموعة الأولى هو 100 وتراوح أعمارهم بين 4 و7 سنوات. ليس لدى أي طفل أي اضطراب لغوي. وأعيد اللقاء مع 10 أطفال بعد مرور سنتين لدراسة تطور معرفتهم للمفردات، وقدرهم على سرد القصص.

المجموعة الثانية مؤلفة من 100 طفل (ليس لديهم اضطراب لغوي) يعيشون في لبنان وتراوح أعمارهم بين 4 و7 سنوات. يتكلمون اللغة العربية بالإضافة إلى الإنجليزية أو الفرنسية. استُعمِلوا معرفتهم المفردات في اللغة العربية فقط.

لأن هذا البحث يتضمن ثلاث اهداف.

الهدف الأول: اكتشاف مدى مهارة الأطفال ثنائي اللغة الناطقين باللغة العربية والسويدية بفهم المفردات وإنتاجها كما القدرة على سرد القصص في كلتا لغتيهما. جميع معلومات (من خلال استبان الولددين) لمعرفة مدى تأثير العوامل الأساسية على معرفة المفردات، وذلك بالسؤال عن مدى استخدام اللغة في المنزل وخارجها. المَكَّمَم المَسْمَوِى من اللغات
الهدف الثاني: تحليل كيفية تطور معرفة المفردات، وقدرة السرد بعد مرور الوقت. اختير عشرة أطفال من مجموعة الأطفال الأربعة سنوات، وأُعيد امتحانهم بعد سنتين لمناقشة معرفتهم المفردات وسرد القصص. طُلب من الأهل تعبئة استمارة استبيانية ثانية للإجابة عن بعض الأسئلة حول تغير استخدام اللغة عند الطفل، وذلك لمناقشة كيفية تأثير العوامل الأساسية على التطور اللغوي.

الهدف الثالث: البحث عن كيفية تأثير العوامل الأساسية على اللغة العربية عند الأطفال ثنائي اللغة الناطقين بالعربية الذين يعيشون في مجتمعين مختلفين: في السودان، حيث اللغة العربية هي لغة الأقليات، بالمقابل في لبنان، حيث اللغة العربية لغة الأكثريّة.

یطرح هذا البحث الأسئلة التالية:

- السؤال الأول: ما مدى معرفة المفردات وقدرة السرد عند الأطفال (عمر 4-7 سنوات) الناطقين باللغة العربية والسويديّة في لغتهم، وما العوامل الأساسية التي تؤثر بشكل ملموس على هذه المعرفة وهذه القدرة؟
- السؤال الثاني: كيف تتطور معرفة المفردات وقدرة السرد عند الأطفال الناطقين بالعربية والسويديّة من سن 4 إلى 6 سنوات في لغتهم؟
- السؤال الثالث: كيف تؤثر العوامل الأساسية على تطور المفردات في اللغة العربية للأطفال ثنائي اللغة الناطقين بالعربية في لبنان، مقارنة بالأطفال الناطقين بالعربية في السويد؟

أجريت ثلاث دراسات للإجابة عن أسئلة البحث تعرض ادناها:

الدراسة الأولى (دراسة مقطعيّة cross-sectional study) اخْتَبِر 100 طفل (بين عمر 4 و 7 سنوات) في معرفتهم مفردات باللغة العربية وألما السويديّة باستخدام أداة المفردات عبر اللغات ال Cross-linguistic Lexical Task, CLT. وَتَخْتَبِر هذه الأداة الأطفال عن مدى معرفتهم فيم وإنتاج الأسماء والأفعال. استُخدمت أداة تقييم متعددة اللغات للحكايات أي MULTilingual Assessment Instrument for Narratives, MAIN (Gagarina et al., 2019; 2015) 2015) لاختبار قدرة الأطفال على سرد القصص. يروي الطفل من خلالها قصصين في كل لغة.
أعيد اختبار 10 أطفال بعد مرور سنتين، أي عند بلوغهم الستة سنوات. وطلب من الأطفال أن يقوموا بنفس التمارين من خلال الأداة المذكورة أعلاه، وذلك لمعرفة مدى تطور معروفتهم للمفردات وتتطور قدرهم على سرد القصص. وأعيد طرح بعض الأسئلة على والدي الأطفال عن مدى سماع الطفل اللغات التي يعرفها واستخدامها.

الدراسة الثالثة (دراسة مقطعيّة)

اختبر 100 طفل (بين عمر 4 و 7 سنوات) لمعروفتهم مفردات باللغة العربية، وذلك باستخدام النسخة العربية من أداة المفردات عبر اللغات (CLT). واستخدم استبيان للوالدين لجمع معلومات حول تأثير العوامل الخلفية على تطور المفردات في اللغة العربية عند الأطفال.

تعرض نتائج الدراسات أدناه بحسب ترتيب وجودتها في كل فصل.

الفصل السادس: دراسة مقطعيّة في السويد:

اختبر 100 طفل ثاني اللغة عن مدى فهمهم المفردات وإنتاجهم في اللغة العربية واللغة السويدية. لم يكن هناك أي فرق بين اللغتين على المستوى العام، بل تساوت النقاط بين اللغتين. كما هو متوقع، سجل الأطفال نقاطاً أعلى في فهم المفردات مقارنة بإنتاج المفردات. تفاوتت درجة إنتاج المفردات أكثر من درجة فهمها في اللغتين. كان التطور في فهم المفردات وإنتاجها بشكل ملحوظ في اللغتين، وبالخصوص بين عمر الأربع والسبع سنوات. حقق العديد من الأطفال أعلى النقاط في فهم المفردات في اللغة السويدية، ولكن لم يفعل ذلك أي طفل في فهم المفردات في اللغة العربية. بالإضافة إلى ذلك، تفاوتت نقاط فهم المفردات في اللغة العربية بدرجة أقل من نقاط الفهم في اللغة السويدية.

ما مدى تأثير العوامل الخلفية على فهم المفردات وإنتاجها؟

أظهر هذا البحث تأثيراً إيجابياً ملحوظاً لتكليف الأهل مع أطفالهم بلغة الأم (أي اللغة العربية) على كل من فهم المفردات وإنتاجها في اللغة العربية، ولم يؤثر سلباً على فهم المفردات وإنتاجها في اللغة السويدية. سجل الأطفال الذين يقرؤون القصص مع أهلهم نقاطاً أعلى في إنتاج المفردات في اللغة العربية من الأطفال الذين لا يقرؤن القصص مع أهلهم. النسبة المرتفعة لسماع الطفل اللغة السويدية خلال يومه أثرت على إنتاج المفردات في اللغة العربية، ولكن ليس على فهمها.

أما اللغة السويدية، فكان هناك تأثير ملحوظ في فهم المفردات وإنتاجها بين الأطفال الذين بدأوا سماع اللغة السويدية بانتظام قبل عمر الثلاث سنوات، مقارنة بما فوق هذا العمر.
السنين. كما أن هناك تأثير سلبي لسماع الأطفال اللغة العربية بشكل كثيف خلال
يومهم على فهم المفردات وإنتاجها في اللغة السويدية.

لم يؤثر المستوى التعليمي للأهل على فهم المفردات وإنتاجها في كل من اللغة العربية واللغة السويدية. يمكن الإطلاع على تفاصيل نتائج البحث وتحليل هذه النتائج في الدراسة.

الفصل السابع، دراسة مقاطعية في السويد: السرد
سرة الأطفال المئة في السويد قصتين من كل من لغتهم وأجابوا على عشرة أسئلة
 المتعلقة بكل قصة وذلك لدراسة تبعية ومدى فهمهم لهذه القصص. سجل الأطفال نقاطاً أعلى في فهم القصص مقارنة بإنجاحها. هذه النتيجة متوقعة وتظهر أن الفهم السريدي يزيد في التطور عمّا هو في الإنتاج السردي وأن الأطفال الأصغر سنًا يفهمون أكثر مما يرون. كان هناك فرق في تطور الفهم والإنتاج في كلتا اللغتين على مستوى المجموعة بأكملها حسب تقدم العمر. ورد استثناء واحد هو في الإنتاج في إحدى القصص باللغة العربية فقط حيث سجل الأطفال الأصغر سنًا نقاطاً تساوي أو نسب مرتفعة
كالأطفال الأكثر سنًا. هذه النتيجة تشير أن نوع القصة تأثير في إنتاج القصص، كما
أ Lehr

كانت هناك سهولة أكبر في فهم "هدف" و"حالة (شعور)" أبطال القصص من
إنتاجها. عُبر عن "هدف" بشكل أكثر عند الأطفال الذين سنًا، مما يعني أن قدرة
إنتاج حلقات تكاملية لأحداث القصة (أي الإبراز عن هدف البطل ومحاولة التي قام
بها لإنجاز هذا الهدف بالإضافة إلى نتيجة هذه المحاولة) كان يَنتج أكثر من الأطفال
الأكبر سنًا. يُظهر هذا البحث أن الأطفال في السنة من عمرهم يمكنهم إنتاج "محاولة"
و"نتيجة المحاولة" بسهولة، ولكنه إنتاج "الهدف" كان ما زال صعباً لأكثرية الأطفال
بعكس فهمهم.

الفصل الثامن، دراسة طويلة في السويد: مفردات و سرد
ظهر تطور نمو معرفة المفردات وقدرة سرد القصص عند الأطفال العشرة (الذين أعيد
اللقاء معهم بعد ستيني) متناسقاً بشكل عام للفرق بين علامات الأطفال ذوي الرأية
من العمر وذويه السادسة من العمر في الدراسة المقاطعية في السويد. ظاهر تطور ملحوظ
عن كل الأطفال في إنتاج وفهم كل من المفردات وسرد القصص في اللغة السويدية
(هناك بعض الاستثناءات البسيطة). لم يكن هذا التطور بنفس القوة لدى جميع الأطفال
من ناحية إنتاج المفردات وفهمها، ومن ناحية سرد القصص وفهمها في اللغة العربية،
حيث سجل بعض الأطفال تطورًا ملحوظًا، بينما البعض الآخر، سجل علامات
متطابقة أو أقل من تلك التي أجروها في عمر الرابعة.
الفصل التاسع، دراسة مقطعية في لبنان: مفردات

اختبر في هذا الفصل مدى معرفة 100 طفل من ثنائيي اللغة فهم المفردات وإنتاجها في اللغة العربية، وقررنت نتائجهما ومدى تأثير العوامل الأساسية عليها مع الأطفال الذين شاركوا في الدراسة المقطعية في السويد. كما هو متوقع، كانت نقاط فهم المفردات أعلى من نقاط الإنتاج. ارتفعت نقاط الفهم والإنتاج بشكل أقوى مع تقدم العمر في لبنان مقارنة بالسويد. في كل البلدين، ارتفعت نقاط الفهم بشكل أقوى مع تقدم العمر من نقاط الإنتاج. قام الأطفال في الدراسة اللبنانية بالتبديل بين لغتيهم (العربية واللغة الثانية أي الإنجليزية أو الفرنسية) بدرجة أكبر من الأطفال في السويد، خاصةً في إنتاج الأسماء.

ما مدى تأثير العوامل الأساسية على فهم المفردات وإنتاجها في اللغة العربية؟ كان هناك تأثير إيجابي لتكلم الأهل اللغة العربية مع أطفالهم على إنتاج المفردات. كما في الدراسة السويدية، لم يكن هناك أي تأثير ملحوظ للمستوى العلمي عند الأهل على فهم المفردات وإنتاجها في اللغة العربية. يصبح بقراءة البحث للمزيد من المعلومات حول نتائج الدراسة.

ختاماً، معرفة المفردات في اللغة السويدية ستكتسب تدريجياً مع العمر من خلال التحاق الطفل بالحضانة أو المدرسة و إخراجه في المجتمع. من الممكن أن تتطور لغته الأم بشكل أبطأ مع مرور العمر ولكن تكلم الأهل مع طفليهم باللغة العربية له تأثير إيجابي ملحوظ. تؤثر العوامل الأساسية على فهم المفردات وإنتاجها، ومن الممكن أن تكون نسبة هذا التأثير مختلفة بين مجتمع وآخر وبين طفل وآخر.

يفهم الطفل أكثر مما ينطق وذلك من ناحية معرفته للمفردات وسرده القصص. يستطيع الطفل ذو عمر السابعة أن ينتج حلقات متسلسلة في القصص أكثر من الأطفال الأصغر سنًا ولكن سرده للفصص لا يشبه إنتاج القصص عند الراغبين. يعزز هذا البحث مفهومنا بتطور اللغة عند الأطفال ثنائيي اللغة في السويد ولبنان، ويشتت آفاقًا جديدة للبحث اللغوي في مجال اكتساب اللغة.
Appendix

Appendix 1 (Chapter 4, Methods)

A1.1 Counterbalancing tables

Table A1.1. Counterbalancing for children first tested in Arabic in the cross-sectional study in Sweden.

<table>
<thead>
<tr>
<th>Child</th>
<th>Arabic 1</th>
<th>Swedish 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NWR</td>
<td>MAIN 1</td>
</tr>
<tr>
<td>1</td>
<td>NWR-Leb</td>
<td>Cat 1</td>
</tr>
<tr>
<td>2</td>
<td>QU-Ara</td>
<td>Cat 2</td>
</tr>
<tr>
<td>3</td>
<td>NWR-Leb</td>
<td>Dog 3</td>
</tr>
<tr>
<td>4</td>
<td>QU-Ara</td>
<td>Dog 4</td>
</tr>
<tr>
<td>5</td>
<td>NWR-Leb</td>
<td>Cat 1</td>
</tr>
<tr>
<td>6</td>
<td>QU-Ara</td>
<td>Cat 2</td>
</tr>
<tr>
<td>7</td>
<td>NWR-Leb</td>
<td>Dog 3</td>
</tr>
<tr>
<td>8</td>
<td>QU-Ara</td>
<td>Dog 4</td>
</tr>
</tbody>
</table>

Table A1.2. Counterbalancing for children first tested in Swedish in the cross-sectional study in Sweden.

<table>
<thead>
<tr>
<th>Child</th>
<th>Swedish 1</th>
<th>Child 1</th>
<th>NWR MAIN 1</th>
<th>CLT MAIN 2</th>
<th>NWR MAIN 1</th>
<th>CLT MAIN 2</th>
<th>NWR</th>
<th>Arabic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>QU-Swe</td>
<td>Cat</td>
<td>1</td>
<td>BB</td>
<td>LS-Swe</td>
<td>QU-Ara</td>
<td>Dog</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>LS-Swe</td>
<td>Cat</td>
<td>2</td>
<td>BG</td>
<td>QU-Swe</td>
<td>NWR-Leb</td>
<td>Dog</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>QU-Swe</td>
<td>Dog</td>
<td>3</td>
<td>BB</td>
<td>LS-Swe</td>
<td>QU-Ara</td>
<td>Cat</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>LS-Swe</td>
<td>Dog</td>
<td>4</td>
<td>BG</td>
<td>QU-Swe</td>
<td>NWR-Leb</td>
<td>Cat</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>QU-Swe</td>
<td>Cat</td>
<td>1</td>
<td>BB</td>
<td>LS-Swe</td>
<td>QU-Ara</td>
<td>Dog</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>LS-Swe</td>
<td>Cat</td>
<td>2</td>
<td>BG</td>
<td>QU-Swe</td>
<td>NWR-Leb</td>
<td>Dog</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>QU-Swe</td>
<td>Dog</td>
<td>3</td>
<td>BB</td>
<td>LS-Swe</td>
<td>QU-Ara</td>
<td>Cat</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>LS-Swe</td>
<td>Dog</td>
<td>4</td>
<td>BG</td>
<td>QU-Swe</td>
<td>NWR-Leb</td>
<td>Cat</td>
<td>4</td>
</tr>
</tbody>
</table>


A1.2 Parental education (SES)

Table A1.3. Highest education achieved by parents of the children in the cross-sectional study in Sweden.

<table>
<thead>
<tr>
<th>ISCED Levels</th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (early childhood education)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1 (primary education)</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>2 (lower secondary education)</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>3 (upper secondary education)</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>4 (post-secondary non-tertiary education)</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>5 (short-cycle tertiary education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 (Bachelor’s studies)</td>
<td>40</td>
<td>34</td>
</tr>
<tr>
<td>7 (Master’s studies)</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8 (Doctoral studies)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Missing information</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table A1.4. Highest education achieved by parents of the children in the longitudinal study in Sweden.

<table>
<thead>
<tr>
<th>ISCED Levels</th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (early childhood education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 (primary education)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2 (lower secondary education)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3 (upper secondary education)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4 (post-secondary non-tertiary education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5 (short-cycle tertiary education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 (Bachelor’s studies)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>7 (Master’s studies)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8 (Doctoral studies)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing information</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table A1.5. Highest education achieved by parents of the children in the cross-sectional study in Lebanon.

<table>
<thead>
<tr>
<th>ISCED Levels</th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (early childhood education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 (primary education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 (lower secondary education)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3 (upper secondary education)</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>4 (post-secondary non-tertiary education)</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>5 (short-cycle tertiary education)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 (Bachelor’s studies)</td>
<td>47</td>
<td>41</td>
</tr>
<tr>
<td>7 (Master’s studies)</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>8 (Doctoral studies)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Missing information</td>
<td>23</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
A1.3 Transcriptions (The cross-sectional study in Sweden: narrative macrostructure)

Examples of narratives in Arabic and in Swedish of low-scoring children and high-scoring children with added English translations by the author. The Arabic transcriptions are presented in the Appendix based on how they were transcribed in the original transcription files and not according to Eid et al. (2011) as in the main text of this dissertation. Thus, ‘2’, ‘3’ and ‘7’ represent /ʾ/, /ʿ/ and /ḥ/, respectively, and ‘sh’ is /š/, ‘kh’ is /ḫ/, ‘j’ is /ǧ/. Swedish transcriptions are presented in standard Swedish orthography.

A1.3.1 Transcription examples in Arabic
BiAra6-11 Arabic MAIN 1 (Dog), narrative macrostructure score = 1 point [low]
@Begin
@Languages: ara
@Participants: EXP Rima Haddad Experimenter, CHI BiAra6-11 Target_Child
@ID: ara|BiLITAS|EXP||female|||Experimenter||
@ID: ara|BiLITAS|CHI|6;4.|female|||Target_Child||
@Date: 4-NOV-2018
@Comment: MAIN1, Dog
@Transcriber: Pascale Wehbe
*EXP: khabrini elossa.
%eng: tell me the story
*CHI: ma a3raf.
%eng: I don’t know
*EXP: mbala, 7awli [/] 7awli.
%eng: sure you do, try, try
*CHI: &eh (...) ma a3raf aysh aqul.
%eng: &eh (...) I don’t know what to say
*EXP: mbala aysh tshufin bi elossa?
%eng: sure you do, what do you see in the story?
*CHI: &eh (...) hay ashu bi el3arabi?
%eng: &eh (...) what is this (called) in Arabic?
%com: CHI points at mouse in pic1.
*EXP: ma a3rif, ana mish, ma 3am shuf, inti [/] inti samiha.
%eng: I don’t know, I’m not, I cannot see, you [/] you name it
*CHI: &eh mus [@s].
%eng: &eh mouse (in Swedish)
*EXP: ok.
*CHI: &eh elawaw w elmus [@s] (...) springa [@s].
%eng: &eh the woof woof and the mouse (in Swedish) (...) run (in Swedish).
%com: the child names the dog as awaw (also uttered as 3aw3aw) which is a common onomatopoetic word (of the barking of the dog) used to indicate the whole dog. The expression is mainly used for and by younger children.
*EXP: ok.
*CHI: jaga [@s] (..), wa7id walid yiji, baloni waka3 fu2 (..) w (...) ba3dan (...) kan ma yjibu (.), awaw w mus [@s] kanu (.) xx.
%eng: chase (in Swedish) (..), a boy comes, balloon fell up, (..) and (…) then (…) he was bringing it (.), woof woof and mouse (in Swedish) were (.) (unintelligible word(s)).
*EXP: ok.
%com: EXP unfolds pics3-4.
and then became (unintelligible word(s)) and then on up, he started to blow (in Swedish) and then woof woof fell mouse (in Swedish) and then he became angry (.) the he woof woof became angry at her mouse (in Swedish) (.) and (.) then they are making (.) fighting (in Swedish) (.).

then he went &aja he brought him, he became (..) happy (in Swedish)

ok.

this is all

hay kil elossa?

is this all?

eh.

yes

sar ba3id shi, 7asal ba3id shi?

did anything else happen?

la2.

no

ok, tayib shukran.

ok, fine thank you

BiAra5-11 Arabic MAIN1 (Dog), narrative macrostructure score = 10 points [high]

@Begin
@Languages: ara
@Participants: EXP Rima Haddad Experimenter, CHI BiAra5-11 Target_Child
@ID: ara|BiLITAS|EXP||female|||Experimenter||
@ID: ara|BiLITAS|CHI|5;6.|female|||Target_Child||
@Date: 31-OCT-2018
@Comment: MAIN1, Dog
@Transcriber: Pascale Wehbe

the story starts here, tell me the best story

(.) once there was a dog (he) wanted to take the mouse

&elbal on the balloon

so he ran (. ) and a man came with a bag and was holding a balloon

&elbal fi elbalon.

&elshaj ara7 fi jabu, sar(.) glad [@s].

then he went &aja he brought him, he became (. ) happy (in Swedish)

ok.

this is all

hay kil elossa?

is this all?

eh.

yes

sar ba3id shi, 7asal ba3id shi?

did anything else happen?

la2.

no

ok, tayib shukran.

ok, fine thank you

hon bitbalish elossa, khabrini a7la ossa.

the story starts here, tell me the best story

( ) marra kan fi kalib biddu yakhud elfar.

did anything else happen?

la2.

no

ok, tayib shukran.

ok, fine thank you

@End
A1.3.2 Transcription examples in Swedish
BiAra5-07 Swedish MAIN2 (Baby Goats), narrative macrostructure score = 1 point [low]
@Begin
@Languages: swe
@Participants: EXP Linnéa_Öberg Experimenter, CHI BiAra5-07 Target_Child
@ID: swe|BiLITAS|EXP||Female|||Experimenter||
@ID: swe|BiLITAS|CHI|5;0.|Male|||Target_Child||
@Date: 08-MAR-2019
@Comment: MAIN2, Baby Goats
@Transcriber: Frauke Jonsson
*EXP: nu får du berätta din saga för mig.
  %eng: you may now tell me your story
  %com: EXP folds back pics so that only pic 1-2 are visible.
*CHI: &e vargen är hungrig för att äta &hm (..).
  %eng: &e the wolf is hungry in order to eat &hm (..)
  %com: CHI points indistinctly to pics from a distance.
*EXP: mhm.
  %com: EXP unfolds pic 3-4.
*CHI: sen han gatt [?] och () hållde [: höll] ben () &ehm () &maa.
  %eng: then he went [?] and () held leg () &ehm () &maa
*EXP: mhm.
*CHI: klart () sen kom det () ett fågel och han slog vargen.
  %eng: finished () then it came () a bird and he hit the wolf
*EXP: mhm.
*CHI: vilken fin saga, jättetbra.
  %eng: what a nice story, very good
@End
EXP: mhm, nu får du berätta din saga för mig.
%eng: mhm, you may now tell me your story
%com: EXP folds back pics so that only pic1-2 are visible.

*CHI: det var tre lamm, &mam &mn stora bocken, han såg den lilla bocken på vatt
net, <den &s> [///] den &e ramrade i vattnet och sen bocken, den stora
bocken, hjälpte den uppe (.)<sen en &törst> [//] sen en törstig räv villde [:
ville] äta upp den.
%eng: there were three sheep, &mam &mn big goat, he saw the little goat on the
water, <the &s> [///] the &e fell in the water and then the goat, the big goat,
helped it up (.)<then a &törst> [//] then a thirsty fox wanted to eat this one
up
%com: CHI points to fox ("sen en törstig räv"), then to grazing baby goat ("villde äta
upp den") in pic2.

*EXP: mhm.
%com: EXP nods; EXP unfolds pic3-4.

*CHI: sen räven hoppade, tilll lilla lammcn, till lilla bocken, <och han tog hans &f
> [/] och den tog hennes fot och sen (.<ett [ ] <ett fägel komde [: kom], och
han tittade på dom.
%eng: then the fox jumped, to the little lamb, to the little goat, <and he took his &f>
[///] and it took her foot and then (.<ett [ ] <ett a bird came, and he looked at them

*EXP: mhm.
*CHI: xx, och &sdom [/] dom [/] dom två bockarna, mellanbocken och stora
bocken, <dom &ä> [/] dom dricker vattnet [=? vattnet].
%eng: (unintelligible word(s)) and &sdom [/] they [/] they the two goats, the middle
goat and the big goat, <they &ä> [/] they drink water
%com: xx: a very short sound, could be some self-affirming interjection ("ja", "ha",
etc.).

*EXP: mhm.
%com: EXP nods; EXP unfolds pic5-6.

*CHI: <sen &s> [/] (.<sen fägel komde [: kom], sen bitit räven, alla tittade på den
och sen alla var glada, fäglen [: fägel] blev arg och räven var ledsen.
%eng: <then &s> [/] (.<then the bird came, then bitten the fox, everyone looked at
it and then everyone was happy, the bird became angry and the fox was sad

*EXP: okej, var det slut så?
%eng: ok, did it finish?
*CHI: ja.
%eng: yes
*EXP: vilken fin saga, bra.
%eng: what a nice story, good

@End
A1.4 MAIN Comprehension Questions

A1.4.1 MAIN1 Cat

Table A1.6. The ten comprehension questions for Cat (MAIN 1).

<table>
<thead>
<tr>
<th>Question</th>
<th>Cat</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Episode 1 Goal</td>
<td>Why does the cat jump/leap forward? <em>(point to pictures 1-2)</em></td>
</tr>
<tr>
<td>D2. Episode 1 IST</td>
<td>How does the cat feel? <em>(point to picture 3)</em></td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
<td>Why do you think that the cat is feeling [answer of D2]?</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
<td>Why does the boy hold the fishing rod in the water? <em>(point to picture 5)</em></td>
</tr>
<tr>
<td>D5. Episode 2 IST</td>
<td>How does the boy feel? <em>(point to picture 6)</em></td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
<td>Why do you think that the boy is feeling [answer of D5]?</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
<td>Why does the cat grab the fish? <em>(point to picture 5)</em></td>
</tr>
<tr>
<td>D8. Episode 3 IST Theory of Mind</td>
<td>Imagine that the boy sees the cat. How does the boy feel? <em>(point to picture 6)</em></td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale Theory of Mind</td>
<td>Why do you think that the boy feels [answer of D8]?</td>
</tr>
<tr>
<td>D10. Overall plotline</td>
<td>Will the boy be friends with the cat? Why?</td>
</tr>
</tbody>
</table>

*Note. IST = Internal State Term.*

A1.4.2 MAIN1 Dog

Table A1.7. The ten comprehension questions for Dog (MAIN1).

<table>
<thead>
<tr>
<th>Question</th>
<th>Dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Episode 1 Goal</td>
<td>Why does the dog jump/leap forward? <em>(point to pictures 1-2)</em></td>
</tr>
<tr>
<td>D2. Episode 1 IST</td>
<td>How does the dog feel? <em>(point to picture 3)</em></td>
</tr>
<tr>
<td>D3. Episode 1 IST rationale</td>
<td>Why do you think that the dog is feeling [answer of D2]?</td>
</tr>
<tr>
<td>D4. Episode 2 Goal</td>
<td>Why does the boy jump/leap upwards? <em>(point to picture 5)</em></td>
</tr>
<tr>
<td>D5. Episode 2 IST</td>
<td>How does the boy feel? <em>(point to picture 6)</em></td>
</tr>
<tr>
<td>D6. Episode 2 IST rationale</td>
<td>Why do you think that the boy is feeling [answer of D5]?</td>
</tr>
<tr>
<td>D7. Episode 3 Goal</td>
<td>Why does the dog grab the sausages? <em>(point to picture 5)</em></td>
</tr>
<tr>
<td>D8. Episode 3 IST Theory of Mind</td>
<td>Imagine that the boy sees the dog. How does the boy feel? <em>(point to picture 6)</em></td>
</tr>
<tr>
<td>D9. Episode 3 IST rationale Theory of Mind</td>
<td>Why do you think that the boy feels [answer of D8]?</td>
</tr>
<tr>
<td>D10. Overall plotline</td>
<td>Will the boy be friends with the dog? Why?</td>
</tr>
</tbody>
</table>

*Note. IST = Internal State Term.*
A1.4.3 MAIN2 Baby Birds

Table A1.8. The ten comprehension questions for Baby Birds (MAIN2).

<table>
<thead>
<tr>
<th>Question</th>
<th>Baby Birds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D1. Episode 1 Goal/ IST</strong></td>
<td>Why does the mother bird fly away? (point to pictures 1-2)</td>
</tr>
<tr>
<td><strong>D2. Episode 1 IST</strong></td>
<td>How do the baby birds feel? (point to picture 1)</td>
</tr>
<tr>
<td><strong>D3. Episode 1 IST rationale</strong></td>
<td>Why do you think that the baby birds are feeling [answer of D2]?</td>
</tr>
<tr>
<td><strong>D4. Episode 2 Goal</strong></td>
<td>Why does the cat climb the tree? (point to picture 3)</td>
</tr>
<tr>
<td><strong>D5. Episode 2 IST</strong></td>
<td>How does the cat feel? (point to picture 5-6)</td>
</tr>
<tr>
<td><strong>D6. Episode 2 IST rationale</strong></td>
<td>Why do you think that the cat is feeling [answer of D5]?</td>
</tr>
<tr>
<td><strong>D7. Episode 3 Goal</strong></td>
<td>Why does the dog grab the cat’s tail? (point to picture 5)</td>
</tr>
<tr>
<td><strong>D8. Episode 3 IST Theory of Mind</strong></td>
<td>Imagine that the dog sees the birds. How does the dog feel? (point to picture 6)</td>
</tr>
<tr>
<td><strong>D9. Episode 3 IST rationale Theory of Mind</strong></td>
<td>Why do you think that the boy feels [answer of D8]?</td>
</tr>
<tr>
<td><strong>D10. Overall plotline</strong></td>
<td>Who does the mother bird like best, the cat or the dog? Why?</td>
</tr>
</tbody>
</table>

Note. IST = Internal State Term.

A1.4.4 MAIN2 Baby Goats

Table A1.9. The ten comprehension questions for Baby Goats (MAIN2).

<table>
<thead>
<tr>
<th>Question</th>
<th>Baby Goats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D1. Episode 1 Goal/ IST</strong></td>
<td>Why was the mother goat in the water? (point to pictures 1-2)</td>
</tr>
<tr>
<td><strong>D2. Episode 1 IST</strong></td>
<td>How do the baby goat feel? (point to picture 1)</td>
</tr>
<tr>
<td><strong>D3. Episode 1 IST rationale</strong></td>
<td>Why do you think that the baby goat is feeling [answer of D2]?</td>
</tr>
<tr>
<td><strong>D4. Episode 2 Goal</strong></td>
<td>Why does the fox jump/leap forward? (point to picture 3)</td>
</tr>
<tr>
<td><strong>D5. Episode 2 IST</strong></td>
<td>How does the fox feel? (point to picture 5-6)</td>
</tr>
<tr>
<td><strong>D6. Episode 2 IST rationale</strong></td>
<td>Why do you think that the fox is feeling [answer of D5]?</td>
</tr>
<tr>
<td><strong>D7. Episode 3 Goal</strong></td>
<td>Why does the bird bite the fox’s tail? (point to picture 5)</td>
</tr>
<tr>
<td><strong>D8. Episode 3 IST Theory of Mind</strong></td>
<td>Imagine that the bird sees the goats. How does the bird feel? (point to picture 6)</td>
</tr>
<tr>
<td><strong>D9. Episode 3 IST rationale Theory of Mind</strong></td>
<td>Why do you think that the bird is feeling [answer of D8]?</td>
</tr>
<tr>
<td><strong>D10. Overall plotline</strong></td>
<td>Who does the mother goat like best, the fox or the bird? Why?</td>
</tr>
</tbody>
</table>

Note. IST = Internal State Term.
Appendix 2 (Chapter 7, Cross-sectional study in Sweden: Narrative macrostructure)

A2.1 Comprehension scores

**Table A2.1.** Arabic and Swedish narrative comprehension scores for Cat and Dog (MAIN1) and Baby Birds and Baby Goats (MAIN2), all ages combined (4;0–7;11).

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cat</td>
<td>Dog</td>
<td>Baby Birds</td>
<td>Baby Goats</td>
</tr>
<tr>
<td>Arabic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>49</td>
<td>50</td>
<td>52</td>
<td>47</td>
</tr>
<tr>
<td>Mean</td>
<td>6.6</td>
<td>6.2</td>
<td>4.7</td>
<td>4.9</td>
</tr>
<tr>
<td>SD</td>
<td>2.8</td>
<td>2.5</td>
<td>2.1</td>
<td>2.6</td>
</tr>
<tr>
<td>Range</td>
<td>1–10</td>
<td>2–10</td>
<td>0.9–9</td>
<td>0–10</td>
</tr>
<tr>
<td>Swedish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>50</td>
<td>47</td>
<td>51</td>
<td>48</td>
</tr>
<tr>
<td>Mean</td>
<td>6.6</td>
<td>6.8</td>
<td>4.5</td>
<td>5.4</td>
</tr>
<tr>
<td>SD</td>
<td>2.8</td>
<td>2.9</td>
<td>2.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Range</td>
<td>0.6–10</td>
<td>0–10</td>
<td>0–10</td>
<td>0–10</td>
</tr>
</tbody>
</table>

*Note.* Max narrative comprehension score = 10 points.

A2.2 Goals in narrative comprehension

**Table A2.2.** Percentage (%) of goal comprehension in Arabic and Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) for each age group.

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-</td>
<td>5-</td>
<td>6-</td>
<td>7-</td>
</tr>
<tr>
<td>Arabic</td>
<td>years</td>
<td>years</td>
<td>years</td>
<td>years</td>
</tr>
<tr>
<td>N</td>
<td>89</td>
<td>92</td>
<td>92</td>
<td>90</td>
</tr>
<tr>
<td>Swedish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### A2.3 Production scores

**Table A2.3** Arabic and Swedish narrative production scores for Cat and Dog (MAIN1) and Baby Birds and Baby Goats (MAIN2), all ages combined (4;0–7;11).

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cat</td>
<td>Dog</td>
<td>Baby Birds</td>
<td>Baby Goats</td>
</tr>
<tr>
<td><strong>Arabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>47</td>
<td>51</td>
<td>51</td>
<td>47</td>
</tr>
<tr>
<td>Mean</td>
<td>5.3</td>
<td>5.4</td>
<td>5.4</td>
<td>5.5</td>
</tr>
<tr>
<td>SD</td>
<td>2.7</td>
<td>2.5</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Range</td>
<td>0–10</td>
<td>0–10</td>
<td>0–11</td>
<td>0–11</td>
</tr>
<tr>
<td><strong>Swedish</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>49</td>
<td>48</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>Mean</td>
<td>4.5</td>
<td>4.9</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>SD</td>
<td>2.5</td>
<td>2.3</td>
<td>2.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Range</td>
<td>0–12</td>
<td>1–9</td>
<td>0–9</td>
<td>0–11</td>
</tr>
</tbody>
</table>

*Note:* Max narrative production score = 17 points.

### A2.3.1 Macrostructure components

**Table A2.4.** Percentage (%) narrative components produced by each age group in Arabic MAIN1 (Cat/Dog) and Arabic MAIN2 (Baby Birds/Baby Goats).

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4- years</td>
<td>5- years</td>
<td>6- years</td>
<td>7- years</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IST as IE</td>
<td>4.8</td>
<td>12.5</td>
<td>13.8</td>
<td>29.2</td>
</tr>
<tr>
<td>Goal</td>
<td>36.5</td>
<td>26.4</td>
<td>23.0</td>
<td>20.8</td>
</tr>
<tr>
<td>Attempt</td>
<td>38.1</td>
<td>54.2</td>
<td>37.9</td>
<td>52.8</td>
</tr>
<tr>
<td>Outcome</td>
<td>58.7</td>
<td>75.0</td>
<td>72.4</td>
<td>81.9</td>
</tr>
<tr>
<td>IST as R</td>
<td>7.9</td>
<td>4.2</td>
<td>3.4</td>
<td>4.2</td>
</tr>
</tbody>
</table>

*Note:* IST as IE = Internal State Term as Initiating Event, IST as R = Internal State Term as Reaction.

**Table A2.5.** Percentage (%) narrative components produced by each age group in Swedish MAIN1 (Cat/Dog) and Swedish MAIN2 (Baby Birds/Baby Goats).

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4- years</td>
<td>5- years</td>
<td>6- years</td>
<td>7- years</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IST as IE</td>
<td>5.0</td>
<td>12.5</td>
<td>6.9</td>
<td>50.0</td>
</tr>
<tr>
<td>Goal</td>
<td>8.3</td>
<td>18.1</td>
<td>23.0</td>
<td>31.9</td>
</tr>
<tr>
<td>Attempt</td>
<td>15.0</td>
<td>23.6</td>
<td>24.1</td>
<td>23.6</td>
</tr>
<tr>
<td>Outcome</td>
<td>15.0</td>
<td>26.4</td>
<td>41.4</td>
<td>55.6</td>
</tr>
<tr>
<td>IST as R</td>
<td>3.3</td>
<td>9.7</td>
<td>6.9</td>
<td>5.6</td>
</tr>
</tbody>
</table>

*Note:* IST as IE = Internal State Term as Initiating Event, IST as R = Internal State Term as Reaction.
### A2.3.2 Story complexity

**Table A2.6.** Percentage (%) of Arabic MAIN1 (Cat/Dog) and Swedish MAIN2 (Baby Birds/Baby Goats) production of story complexity for each age group.

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-</td>
<td>5-</td>
<td>6-</td>
<td>7-</td>
</tr>
<tr>
<td>No seq.</td>
<td>years</td>
<td>years</td>
<td>years</td>
<td>years</td>
</tr>
<tr>
<td></td>
<td>46.0</td>
<td>38.9</td>
<td>54.0</td>
<td>43.1</td>
</tr>
<tr>
<td>AO</td>
<td>19.0</td>
<td>30.6</td>
<td>20.7</td>
<td>36.1</td>
</tr>
<tr>
<td>GA/GO</td>
<td>20.6</td>
<td>13.9</td>
<td>11.5</td>
<td>11.1</td>
</tr>
<tr>
<td>GAO</td>
<td>14.3</td>
<td>16.7</td>
<td>13.8</td>
<td>9.7</td>
</tr>
</tbody>
</table>

*Note. AO = Attempt Outcome, GA/GO = Goal Attempt/ Goal Outcome, GAO = Goal Attempt Outcome, No seq. = No sequence.*

**Table A2.7.** Percentage (%) of Swedish MAIN1 (Cat/Dog) and Swedish MAIN2 (Baby Birds/Baby Goats) production of story complexity for each age group.

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-</td>
<td>5-</td>
<td>6-</td>
<td>7-</td>
</tr>
<tr>
<td>No seq.</td>
<td>years</td>
<td>years</td>
<td>years</td>
<td>years</td>
</tr>
<tr>
<td></td>
<td>85.0</td>
<td>61.1</td>
<td>55.2</td>
<td>40.3</td>
</tr>
<tr>
<td>AO</td>
<td>5.0</td>
<td>20.8</td>
<td>26.4</td>
<td>37.5</td>
</tr>
<tr>
<td>GA/GO</td>
<td>8.3</td>
<td>12.5</td>
<td>10.3</td>
<td>11.1</td>
</tr>
<tr>
<td>GAO</td>
<td>1.7</td>
<td>5.6</td>
<td>8.0</td>
<td>11.1</td>
</tr>
</tbody>
</table>

*Note. AO = Attempt Outcome, GA/GO = Goal Attempt/ Goal Outcome, GAO = Goal Attempt Outcome, No seq. = No sequence.*
Appendix 3 (Chapter 8, Longitudinal study in Sweden)

A3.1 Macrostructural components

**Table A3.1.** Percentage (%) narrative components produced by 4-year-olds and 6-year-olds in Arabic MAIN1 (Cat/Dog) and Arabic MAIN2 (Baby Birds/Baby Goats) in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th>Setting</th>
<th>MAIN1</th>
<th>MAIN2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-sectional</td>
<td>Longitudinal</td>
</tr>
<tr>
<td>4-years</td>
<td>6-years</td>
<td>4-years</td>
</tr>
<tr>
<td>Setting</td>
<td>4.8</td>
<td>13.8</td>
</tr>
<tr>
<td>IST as IE</td>
<td>7.9</td>
<td>23.0</td>
</tr>
<tr>
<td>Goal</td>
<td>36.5</td>
<td>28.7</td>
</tr>
<tr>
<td>Attempt</td>
<td>38.1</td>
<td>37.9</td>
</tr>
<tr>
<td>Outcome</td>
<td>58.7</td>
<td>72.4</td>
</tr>
<tr>
<td>IST as R</td>
<td>7.9</td>
<td>3.4</td>
</tr>
</tbody>
</table>

*Note. IST as IE = Internal State Term as Initiating Event, IST as R = Internal State Term as Reaction.*

**Table A3.2.** Percentage (%) narrative components produced by 4-year-olds and 6-year-olds in Swedish MAIN1 (Cat/Dog) and Swedish MAIN2 (Baby Birds/Baby Goats) in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th>Setting</th>
<th>MAIN1</th>
<th>MAIN2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-sectional</td>
<td>Longitudinal</td>
</tr>
<tr>
<td>4-years</td>
<td>6-years</td>
<td>4-years</td>
</tr>
<tr>
<td>Setting</td>
<td>5.0</td>
<td>6.9</td>
</tr>
<tr>
<td>IST as IE</td>
<td>8.3</td>
<td>23.0</td>
</tr>
<tr>
<td>Goal</td>
<td>15.0</td>
<td>24.1</td>
</tr>
<tr>
<td>Attempt</td>
<td>15.0</td>
<td>41.4</td>
</tr>
<tr>
<td>Outcome</td>
<td>50.0</td>
<td>60.9</td>
</tr>
<tr>
<td>IST as R</td>
<td>3.3</td>
<td>6.9</td>
</tr>
</tbody>
</table>

*Note. IST as IE = Internal State Term as Initiating Event, IST as R = Internal State Term as Reaction.*
### A3.2 Story complexity

**Table A3.3.** Percentage (%) of Arabic MAIN1 (Cat/Dog) and Arabic MAIN2 (Baby Birds/Baby Goats) production of story complexity produced by 4-year-olds and 6-year-olds in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-sectional</td>
<td>Longitudinal</td>
<td>Cross-sectional</td>
<td>Longitudinal</td>
</tr>
<tr>
<td>4-year</td>
<td>6-year</td>
<td>4-year</td>
<td>6-year</td>
<td>4-year</td>
</tr>
<tr>
<td>No seq.</td>
<td>46.0</td>
<td>54.0</td>
<td>55.6</td>
<td>59.3</td>
</tr>
<tr>
<td>AO</td>
<td>19.0</td>
<td>20.7</td>
<td>11.1</td>
<td>18.5</td>
</tr>
<tr>
<td>GA/GO</td>
<td>20.6</td>
<td>11.5</td>
<td>14.8</td>
<td>7.4</td>
</tr>
<tr>
<td>GAO</td>
<td>14.3</td>
<td>13.8</td>
<td>18.5</td>
<td>14.8</td>
</tr>
</tbody>
</table>

*Note. AO = Attempt Outcome, GA/GO = Goal Attempt/Goal Outcome, GAO = Goal Attempt Outcome, No seq. = No sequence.*

**Table A3.4.** Percentage (%) of Swedish MAIN1 (Cat/Dog) and Swedish MAIN2 (Baby Birds/Baby Goats) production of story complexity produced by 4-year-olds and 6-year-olds in the cross-sectional study and the longitudinal study.

<table>
<thead>
<tr>
<th></th>
<th>MAIN1</th>
<th></th>
<th>MAIN2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross-sectional</td>
<td>Longitudinal</td>
<td>Cross-sectional</td>
<td>Longitudinal</td>
</tr>
<tr>
<td>4-year</td>
<td>6-year</td>
<td>4-year</td>
<td>6-year</td>
<td>4-year</td>
</tr>
<tr>
<td>No seq.</td>
<td>85.0</td>
<td>55.2</td>
<td>88.9</td>
<td>43.3</td>
</tr>
<tr>
<td>AO</td>
<td>5.0</td>
<td>26.4</td>
<td>3.7</td>
<td>33.3</td>
</tr>
<tr>
<td>GA/GO</td>
<td>8.3</td>
<td>10.3</td>
<td>3.7</td>
<td>13.3</td>
</tr>
<tr>
<td>GAO</td>
<td>1.7</td>
<td>8.0</td>
<td>3.7</td>
<td>10.0</td>
</tr>
</tbody>
</table>

*Note. AO = Attempt Outcome, GA/GO = Goal Attempt/Goal Outcome, GAO = Goal Attempt Outcome, No seq. = No sequence.*
Appendix 4 (Chapter 9, Cross-sectional study in Lebanon: Vocabulary)

A4.1 Arabic vocabulary comprehension and production scores in Lebanon and Sweden

**Table A4.1.** Arabic vocabulary (CLT) comprehension and production scores for each age group in Lebanon and in Sweden.

<table>
<thead>
<tr>
<th></th>
<th>Lebanon</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>5-year-olds</td>
<td>6-year-olds</td>
<td>7-year-olds</td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>29</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>44.39</td>
<td>48.34</td>
<td>50.87</td>
<td>55.41</td>
</tr>
<tr>
<td>SD</td>
<td>5.99</td>
<td>5.09</td>
<td>5.13</td>
<td>2.61</td>
</tr>
<tr>
<td>Range</td>
<td>31–52</td>
<td>36–55</td>
<td>32–59</td>
<td>51–60</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>24.28</td>
<td>27.07</td>
<td>30.84</td>
<td>41.55</td>
</tr>
<tr>
<td>SD</td>
<td>8.53</td>
<td>7.82</td>
<td>9.60</td>
<td>7.47</td>
</tr>
<tr>
<td>Range</td>
<td>10–37</td>
<td>10–43</td>
<td>9–45</td>
<td>28–56</td>
</tr>
<tr>
<td>Sweden</td>
<td>4-year-olds</td>
<td>5-year-olds</td>
<td>6-year-olds</td>
<td>7-year-olds</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td>25</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>41.55</td>
<td>46.72</td>
<td>48.52</td>
<td>52.42</td>
</tr>
<tr>
<td>SD</td>
<td>7.64</td>
<td>6.48</td>
<td>7.51</td>
<td>3.62</td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25.55</td>
<td>32.64</td>
<td>34.48</td>
<td>37.13</td>
</tr>
<tr>
<td>SD</td>
<td>12.15</td>
<td>11.71</td>
<td>13.24</td>
<td>9.21</td>
</tr>
<tr>
<td>Range</td>
<td>1–42</td>
<td>11–48</td>
<td>10–53</td>
<td>16–52</td>
</tr>
</tbody>
</table>

*Note.* Max comprehension score = 60 points, max production score = 60 points.
A4.2 Arabic vocabulary production and conceptual production scores in Lebanon and Sweden

Table A4.2. Arabic vocabulary (CLT) production scores and conceptual production scores for each age group in Lebanon and in Sweden.

<table>
<thead>
<tr>
<th></th>
<th>Lebanon</th>
<th></th>
<th>Sweden</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>5-year-olds</td>
<td>6-year-olds</td>
<td>7-year-olds</td>
</tr>
<tr>
<td></td>
<td>N = 18</td>
<td>N = 29</td>
<td>N = 31</td>
<td>N = 22</td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>24.28</td>
<td>27.07</td>
<td>30.84</td>
<td>41.55</td>
</tr>
<tr>
<td>SD</td>
<td>8.53</td>
<td>7.82</td>
<td>9.60</td>
<td>7.47</td>
</tr>
<tr>
<td>Range</td>
<td>10–37</td>
<td>10–43</td>
<td>9–45</td>
<td>28–56</td>
</tr>
<tr>
<td><strong>Conceptual Production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>36.28</td>
<td>39.38</td>
<td>41.94</td>
<td>47.64</td>
</tr>
<tr>
<td>SD</td>
<td>6.21</td>
<td>5.70</td>
<td>5.60</td>
<td>4.57</td>
</tr>
<tr>
<td>Range</td>
<td>20–44</td>
<td>22–54</td>
<td>25–50</td>
<td>39–56</td>
</tr>
</tbody>
</table>

|              | 4-year-olds      | 5-year-olds  | 6-year-olds      | 7-year-olds  |
|              | N = 22           | N = 25       | N = 29           | N = 24       |
| **Production** |                  |              |                  |              |
| Mean         | 25.55            | 32.64        | 34.48            | 37.13        |
| SD           | 12.15            | 11.71        | 13.24            | 9.21         |
| Range        | 1–42             | 11–48        | 10–53            | 16–52        |

<table>
<thead>
<tr>
<th></th>
<th>Arabic comprehension</th>
<th>Arabic production</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AoO to Arabic</strong></td>
<td>From birth (N = 72)</td>
<td>After age 1 (N = 16)</td>
</tr>
<tr>
<td>Mean</td>
<td>51.28</td>
<td>45.56</td>
</tr>
<tr>
<td>SD</td>
<td>4.71</td>
<td>6.21</td>
</tr>
<tr>
<td>Range</td>
<td>37–60</td>
<td>34–56</td>
</tr>
<tr>
<td><strong>AoO to Arabic</strong></td>
<td>From birth (N = 72)</td>
<td>After age 1 (N = 16)</td>
</tr>
<tr>
<td>Mean</td>
<td>32.81</td>
<td>21.75</td>
</tr>
<tr>
<td>SD</td>
<td>9.61</td>
<td>8.68</td>
</tr>
<tr>
<td>Range</td>
<td>10–56</td>
<td>9–41</td>
</tr>
</tbody>
</table>

*Note. Max production score = 60 points.*

A4.3 Age of onset (AoO) of Arabic and a second language

Table A4.3. Children’s AoO to Arabic (from birth / after age 1).

<table>
<thead>
<tr>
<th></th>
<th>Lebanese</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AoO to Arabic</td>
<td>AoO to Arabic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>From birth (N = 72)</td>
<td>After age 1 (N = 16)</td>
<td></td>
</tr>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>51.28</td>
<td>45.56</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>4.71</td>
<td>6.21</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>37–60</td>
<td>34–56</td>
<td></td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32.81</td>
<td>21.75</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>9.61</td>
<td>8.68</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>10–56</td>
<td>9–41</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Max comprehension score = 60 points, max production score = 60 points.*
Table A4.4. Children’s AoO to English/French (before age 3 / starting from 3).

<table>
<thead>
<tr>
<th>Lebanon</th>
<th>AoO to English/French Before age 3 (N = 77)</th>
<th>AoO to Arabic Starting from 3 (N = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>SD</td>
<td>5.36</td>
<td>5.87</td>
</tr>
<tr>
<td>Range</td>
<td>34–60</td>
<td>37–59</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.0</td>
<td>32.45</td>
</tr>
<tr>
<td>SD</td>
<td>10.14</td>
<td>10.24</td>
</tr>
<tr>
<td>Range</td>
<td>9–56</td>
<td>10–45</td>
</tr>
</tbody>
</table>

*Note. Max comprehension score = 60 points, max production score = 60 points.*

Table A4.5. Children’s AoO to English/French (from birth / after age 1).

<table>
<thead>
<tr>
<th>Lebanon</th>
<th>AoO to English/French From birth (N = 38)</th>
<th>AoO to English/French After age 1 (N = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>49.92</td>
<td>50.48</td>
</tr>
<tr>
<td>SD</td>
<td>6.02</td>
<td>5.01</td>
</tr>
<tr>
<td>Range</td>
<td>34–59</td>
<td>36–60</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>28.34</td>
<td>32.66</td>
</tr>
<tr>
<td>SD</td>
<td>11.26</td>
<td>9.25</td>
</tr>
<tr>
<td>Range</td>
<td>9–56</td>
<td>10–51</td>
</tr>
</tbody>
</table>

*Note. Max comprehension score = 60 points, max production score = 60 points.*

A4.4 Children’s language use siblings

Table A4.6. Child’s language use (only/mostly Arabic vs. other) with siblings in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th>Lebanon</th>
<th>Only/mostly Arabic (N = 21)</th>
<th>Other (N = 58)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>50.6</td>
<td>49.8</td>
</tr>
<tr>
<td>SD</td>
<td>5.5</td>
<td>5.6</td>
</tr>
<tr>
<td>Range</td>
<td>37–60</td>
<td>34–58</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32.2</td>
<td>29.9</td>
</tr>
<tr>
<td>SD</td>
<td>9.5</td>
<td>10.4</td>
</tr>
<tr>
<td>Range</td>
<td>10–50</td>
<td>9–56</td>
</tr>
</tbody>
</table>

*Note. Max comprehension score = 60 points, max production score = 60 points.*
Table A4.7. Child’s language use (only/mostly a second language vs. other) with siblings in relation to Arabic vocabulary (CLT) scores.

<table>
<thead>
<tr>
<th>Lebanon</th>
<th>Only/mostly a second language (N = 17)</th>
<th>Other (N = 62)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>50.7</td>
</tr>
<tr>
<td>SD</td>
<td>6.3</td>
<td>5.2</td>
</tr>
<tr>
<td>Range</td>
<td>34–57</td>
<td>36–60</td>
</tr>
<tr>
<td>Arabic production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>26.9</td>
<td>31.5</td>
</tr>
<tr>
<td>SD</td>
<td>12.5</td>
<td>9.3</td>
</tr>
<tr>
<td>Range</td>
<td>9–56</td>
<td>10–50</td>
</tr>
</tbody>
</table>

*Note.* Max comprehension score = 60 points, max production score = 60 points.


National Agency for Education (Skolverket). (2017). *Antal elever i grundskolan som det finns information om att de är berättigade till och deltar i modersmåls undervisning i arabiska [Number of pupils in grade 1–9 eligible for and participating in mother tongue instruction in Arabic]*. https://www.skolverket.se/polopolyfs/1.259231/


