ACTA UNIVERSITATIS UPSALIENSIS
Studia Linguistica Upsaliensia
19

Developing narrative competence

Swedish, Swedish-German and Swedish-Turkish children aged 4–6

Josefin Lindgren

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Ihresalen, Engelska parken, Thunbergsvägen 3H, Uppsala, Saturday, 19 May 2018 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Associate Professor Sharon Unsworth (Radboud University Nijmegen).

Abstract
Lindgren, J. 2018. Developing narrative competence. Swedish, Swedish-German and Swedish-Turkish children aged 4–6. *Studia Linguistica Upsaliensia* 19. 292 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0292-8.

This thesis investigates the development of oral narrative competence from age 4 to 6 in Swedish monolinguals (N=72) and in both languages of Swedish-German (N=46) and Swedish-Turkish (N=48) bilinguals growing up in Sweden. Picture-based fictional narratives were elicited with Cat/Dog and Baby Birds/Baby Goats from the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al. 2012) and A2/B2 from the Edmonton Narrative Norms Instrument (ENNI, Schneider et al., 2005). Vocabulary, character introduction and narrative macrostructure were studied. Vocabulary production scores on Cross-linguistic lexical tasks (CLTs, Haman et al., 2015) were compared to NDW (number of different words) in narratives. Production of macrostructural components, macrostructural complexity, and answers to comprehension questions were analyzed. Effects of age and differences in performance between groups, between the bilinguals' two languages, and between narrative tasks were investigated.

Narrative comprehension was high already at age 4, but still developed substantially with age. In contrast, macrostructure in narrative production was at a rudimentary level at age 4. Even at age 6, the narratives contained few complete episodic structures. Children mainly included actions visible in the stimuli and rarely verbalized goals and other macrostructural components that required inferencing. The ability to introduce story characters appropriately developed strongly from age 4 to 6, but stimuli had a large effect on performance. Vocabulary showed most improvement from age 5 to 6. Development with age was clearer for the majority language Swedish than the minority languages German and Turkish, where individual variation was larger.

In Swedish, pronounced differences were found between the bilingual groups. The Swedish-German bilinguals performed similarly to the monolinguals. On most measures, the Swedish-Turkish bilinguals performed lower than the other two groups, though precisely how much varied across measures. Generally, the Swedish-German children performed better in Swedish than in German, whereas the Swedish-Turkish children performed similarly in both languages or slightly higher in Turkish. The study shows that bilinguals' two languages need not develop in parallel, and that results depend on the tasks and specific measures used. Bilingual groups differ from each other, and it is therefore not meaningful to compare all bilinguals to all monolinguals.

Keywords: narratives, macrostructure, referent introduction, vocabulary, preschool children, bilingualism, Swedish, German, Turkish, MAIN, ENNI, CLT

Josefin Lindgren, Department of Linguistics and Philology, Box 635, Uppsala University, SE-75126 Uppsala, Sweden.

ISSN 1652-1366
ISBN 978-91-513-0292-8
urn:nbn:se:uu:diva-347102 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-347102)

Acknowledgments

These last five years have been interesting, challenging, stimulating, hard, fun, and somewhat stressful and frustrating at times, but also very beautiful. This project would have been more difficult and more boring, and this book would have been less good without all the people surrounding me.

First of all, I am deeply grateful to my main supervisor Ute Bohnacker, for being so generous with your time and expertise, for always giving precise and constructive comments on my work while at the same time helping me discover and learn things on my own, for your dedication to teaching and to good scientific practices, and for even collecting part of the German data for this study yourself! From the first class I took for you in 2007, throughout my BA, MA and now PhD, you have been a source of inspiration to me. Without your support, creating and carrying out this project would have been much more difficult – thank you for everything!

I also want to thank my second supervisor Taina Välimaa, for helping me focus on my research questions during my visit to Oulu and for valuable feedback on the manuscript, and Natalia Ganuza, my 'final seminar' opponent, whose detailed comments helped me improve this work substantially.

I am very happy to have had the privilege of working together with the BiLI-TAS team! Thanks again to Ute for leading our group, and to the other members Sibylle Dillström, Rima Haddad, Karin Koltay, Linnéa Öberg and Buket Öztekin for good collaboration, for all the long, but fun meetings, for fruitful discussions of earlier versions of my chapters and for always being there when I needed to chat about any subject. Thanks also to Buket for transcribing and scoring the Turkish data, to Sibylle for German native-speaker support and transcript checks, to Linnéa for checking the Swedish transcripts, and to Karin for valuable contributions to the macrostructure scoring and for your unwavering positive attitude! Thanks also to the BiLI-TAS research assistants Ercan Aras, Mikaela Farkas-Behndig, Maria Johansson, Zubeyde Küçükgöl and Müjde Nordling. A big *danke schön* to Valerie Reichardt, 'my own assistant', not only for assisting with the German data collection and transcription, but also for all the interesting discussions.

Tack, *danke*, *teşekkürler* to all the children who participated in the study and their families: without you this work would not have been possible! I also want to thank all (pre)schools who allowed us to visit them and collect data from the children, as well as the (pre)school teachers and mother-tongue

teachers who helped recruiting children; *ett särskilt tack till* Brita Alonzo, Catrin Flodén, Angelika Havik, Ingrid Jansson and Lena Lithén.

I am also grateful to Anna Maria Lundins stipendiefond and Stiftelsen Lundborgska Idofonden for their generous grants, which made it possible for me to attend a number of summer schools and conferences.

During my years as a PhD student, I have spent time with a lot of nice people both inside and outside work. Special thanks to Miryam ('best office mate') and Fabienne ('baker of most delicious cookies') for the walks and the talks; to the Williams gang (you know who you are) for taking my mind off the thesis; to Rima, Linnéa, Buket & Karin again for our fun dinners and corridor meetings; to Miriam T for taking the time to read some of my chapters as well as for all the fikas and lunches and for making me an honorary member of the Moderna språk PhD family; to Vera for keeping me company during the first years; to Tiina for long and interesting discussions about methodology, life and everything and for hanging around during weekends at the office; to Constanze for nice collaboration; to the members of our great department choir Lingsing and especially our conductor Karin (again!) for creating such a warm and positive singing environment; to the members of WoGEL (Working Group of Empirical Linguistics) for all the interesting Friday sessions (with fruit!); and to all other friends and colleagues at the Department of Linguistics & Philology (especially current and former PhD students and staff in Linguistics and Computational Linguistics) and at the rest of the Faculty of Languages – you all made working here better! Thanks also to Emanuel and Fredrik for all the visits to the nations and Café Årummet with lots of discussions about all types of language-related topics, and to Anna, Malin, Cosima, Frederik, Sara, Mareen, Katrin, Elena, Baukje, Erika, Maija, Sebastian, Camilla, Christine, Johan, Daniel, and Dennis for being really good friends!

Tack to my parents Rolf and Ylva, för att ni omgett mig med böcker och för att ni alltid uppmuntrat mig att gå min egen väg och göra det jag tror på, and to my sisters Carolin and Matilda and my nephew Ludvig för alla roliga lekar och alla kramar!

Finally, to my partner in everything, Jorrig, *dankjewel voor alles*! It is not easy to quantify the amount of feedback and general support you have given me with this project and how you continuously help me improve my work, but it is even more difficult, if not impossible, to describe what you give me by simply being you. Du är den finaste. We maken elkaar beter en ik ben zo blij dat we bij elkaar zijn.

Uppsala, April 2018
Josefin Lindgren

Contents

1 Introduction

An increasing number of children in Sweden grow up speaking two or more languages, either as simultaneous bilinguals, learning two languages from birth or shortly thereafter, or as early sequential bilinguals, learning one language first, the L1, and the second, the L2, in early childhood. In 2017, at least 24% of the children in Sweden were growing up bilingually (Statistics Sweden, 2017b).[1] However, relatively little is known about bilingual language development in a Swedish context, especially with regards to preschool children. In fact, even internationally, few large-scale studies have been conducted in which both languages of child bilinguals were analyzed. Existing studies are limited to a few language combinations and contexts. No study has been carried out in Sweden that compares the language development of monolinguals and different bilingual groups, analyzing data from the bilinguals' two languages.

Narrative competence, the ability to tell a story, a series of events in a structured way that makes it clear to the listener what happened to whom and why, is an important part of language proficiency. Narrative language functions as a bridge between oral and written language and is important for later school achievements (Dickinson & Tabors, 2001). Still, little is known about how and when Swedish-speaking children learn to tell stories. The current study investigates the development of oral narrative competence in monolingual Swedish, Swedish-German bilingual and Swedish-Turkish bilingual children aged 4–6 growing up in Sweden. Turkish and German, two typologically different languages, are both sizeable minority languages in Sweden (cf. Parkvall, 2015), yet we know very little about the development of these languages in a Swedish context.

Narration is a universal human activity, but to tell a story in a way that fits the cultural context, the story content and the listener is not easy. Narratives have their own rules and structures, including organizational patterns representing temporal and causal information, forming a narrative schema (Berman & Slobin, 1994b; Labov, 1972; Labov & Waletzky, 1967; Stein & Glenn, 1979; Westby, 2012). A prerequisite for telling a (good) story (cf.

[1] There are no official records of which languages Swedish residents speak. Therefore, this figure is based on the number of children aged 0–14, who are living in Sweden and were either born in another country or whose parents were born in another country; children with one Sweden-born parent are not included here, and the real figure may thus be even higher.

Section 1.2) is understanding the listener's perspective and how it may differ from one's own (using Theory of Mind, cf. e.g. Tomasello, 2003). Another one is the ability to verbalize an underlying story schema (including drawing inferences about goals and internal states of protagonists if necessary, cf. Section 1.1). A third prerequisite is a basic level of language proficiency, e.g. in vocabulary and morphosyntax, without which the speaker will not be able to formulate the story content in an understandable manner (cf. Viberg, 2001). Narrative competence becomes increasingly important as children grow older, both socially and for success in the educational system, for example for literacy development and reading comprehension (e.g. Dickinson & Tabors, 2001; Griffin, Hemphill, Camp, & Wolf, 2004; Gutiérrez-Clellen, 2002).

Starting from age 3–4, children's narratives develop from loosely connected utterances into coherent episodic structures (e.g. Berman & Slobin, 1994c), making narrative tasks suitable for children aged 3–4 and older. Narratives "provide information about how well children can use their discrete language skills to communicate" (Schneider, Hayward, & Dubé, 2006, p. 224) and "are a valuable diagnostic tool because they provide information about a child's ability to plan discourse at the extended level" (Fiestas & Peña, 2004, p. 156). Narrative data also allow for many different types of linguistic analysis (e.g. Berman & Slobin, 1994a; Hickmann, 2004; see also the overview in Pavlenko, 2008) and may be a more ecologically valid way to assess a child's language proficiency and development than tests of specific morphological and/or syntactic structures.

A number of studies have analyzed the development of oral narrative competence or ability[2] in monolingual children (e.g. Berman & Slobin, 1994c), but fewer studies exist for bilingual children. Earlier studies have used different types of stimulus material and focused on different, often quite specific, aspects of narratives, which makes it difficult to compare the results. A number of studies have been published on Spanish-English bilingual children (e.g. Fiestas & Peña, 2004; Pearson, 2002), but there are also some studies of other language combinations, such as Dutch-Turkish, French-Turkish and Dutch-Moroccan Arabic bilinguals (see Aarssen, 1996; Akinci, Jisa, & Kern, 2001; Bos, 1997). There are no studies of narratives by Swedish-German or Swedish-Turkish children.

Research on early monolingual acquisition of Swedish has focused mainly on morphology and syntax (e.g. Josefsson, Platzack, & Håkansson, 2003; Waldmann, 2008). There are some studies of bilingual children in Sweden, but only a few language combinations, such as Swedish-Finnish, have been

[2] The terms *competence* and *ability* are used interchangeably throughout this work; a discussion of the difference between these two concepts falls outside the scope of the present study.

studied more extensively.[3] Generally, relatively small groups of children have been studied. Mostly, the focus has been on the development of Swedish; seldom has data from the bilinguals' other language(s) been analyzed (for a recent exception, see Ganuza & Hedman, 2017). In some cases, Swedish data from different groups of bilinguals have been combined and compared to monolinguals. A few longitudinal case studies of simultaneously bilingual children have investigated the development in both languages (e.g. Bernardini & Schlyter, 2004; Granfeldt, 2000). In addition, there are some studies of Swedish-speaking children's narratives (e.g. Nordqvist, 2001; Viberg, 2001). With the exception of a recent study by Bohnacker (2016) of 52 Swedish-English bilingual 5–7-year-olds, no larger study that analyzes data from both languages has been published on Swedish bilingual children's narratives. For these reasons, there is a need for large-scale studies of Swedish-speaking bilinguals.

The current study is unique in many respects. It is the first large-scale study comparing narratives told by Swedish mono- and bilingual children. Additionally, it is one of the first studies of narratives in which different groups of bilinguals speaking the same majority language but different minority languages are compared.[4] Performance in the bilinguals' two languages (Swedish/German, Swedish/Turkish) is compared, making it possible to get a fuller picture of the bilinguals' narrative competence.

This thesis thus constitutes a first step towards mapping the development of bilingual children's narrative abilities in both languages for different language combinations in Sweden. This development is to a large extent unknown. The study contributes to a better understanding of both children's narrative competence and (bilingual) language development more generally. Previous studies of children's narratives have often employed different narrative tasks, making it difficult to generalize results across languages. The current study uses the same narrative tasks with different groups of children and in all three languages (Swedish, German, Turkish). Moreover, the current study also uses different types of narrative tasks with the same group of children, which allows us to broaden our understanding of how the task may influence performance. The data collected in the study form a starting point for a larger corpus of narratives by Swedish monolingual and bilingual children.

[3] Additionally, there are a number of studies of Swedish school-aged children and adolescents acquiring a foreign language in a school context, for example English (e.g. Sundqvist, 2009) or German (e.g. Rosén, 2008).

[4] In the present study, the terms *minority* and *majority* language refer to whether or not a language is the (official) language of a country, spoken by the majority of the population (majority language), i.e. Swedish in Sweden, or a language that is spoken by a smaller group (minority language), e.g. Turkish and German in Sweden. The term minority language as it is used here does thus not entail that the language in question has any type of official minority language status in the country.

This thesis proceeds as follows. The remainder of Chapter 1 provides the background to the project and explains some of the basic methodological choices made. Chapter 2 states the aims and research questions. In Chapter 3, the methodology is described, with detailed sections on the participants (Section 3.1), elicitation instruments (Section 3.2), setup and procedure (Section 3.3) and aspects of transcription (Section 3.4). Chapter 4 gives an overview of the corpus of narrative data collected in the project. Chapters 5 to 8 report the results. Each results chapter also includes an overview of relevant findings from the literature as well as a discussion. Chapter 5 deals with vocabulary, including vocabulary production scores (Section 5.2) and narrative vocabulary (Section 5.3). Chapter 6 reports results for character introduction. Chapter 7 describes the results for macrostructure, for comprehension (Section 7.2), production (Section 7.3), and the comparison of comprehension and production (Section 7.4). The final results chapter (Chapter 8) links vocabulary and character introduction to macrostructure. Chapter 9 summarizes the results of the study. Finally, Chapter 10 contains a general discussion and conclusions.

1.1 What is a narrative?

There are various ways to describe what a narrative (or story) is. Intuitively a narrative describes events that are temporally or causally linked and in which some protagonist performs some action. Narrative models try to capture the components of the global structure of the narrative, i.e. its hierarchical organization, commonly called the *macrostructure*. In this section, widely used narrative models (or schemata) are described, including the one behind the *Multilingual Assessment Instrument for Narratives* (Gagarina et al., 2012, 2015), one of the narrative assessment tools used in the current study (see Section 3.2).

The simplest definition of a narrative is what has been called a *minimal narrative*: "a sequence of two clauses which are *temporally ordered*" (Labov, 1972, p. 360). However, this definition may not be specific enough to distinguish between narratives and other text types such as descriptive or expository texts (cf. Stein & Policastro, 1984). A central aspect of a stretch of narrative discourse is that it cannot be purely descriptive, but must contain at least one (animate) protagonist performing some action (Stein & Policastro, 1984, p. 147). Another basic criterion of a narrative is that there should be a 'beginning', describing the *onset* of the plot, a 'middle' part, which outlines the *unfolding* of the plot, and an 'end', including the *resolution* of the plot. The division of a narrative's global structure into these three *core plot components* is based on the work by Labov & Waletsky (1967) (see also Labov, 1972, pp. 363, 369–370). Core plot components were for example analyzed in the seminal study by Ruth Berman, Dan Slobin and colleagues (Berman & Slobin, 1994c) on narratives elicited with the wordless picture

book *Frog, where are you?* (Mayer, 1969, hereafter *Frog story*) from monolingual children aged 3–9 and adults speaking English, German, Hebrew, Spanish and Turkish (see Section 7.1 for a description of this and other studies of macrostructure).

Narrative content has also been analyzed in terms of the number of information units or propositions, also called thematic units/elements (e.g. Berman, 1988; Hudson & Shapiro, 1991; Mäkinen, 2014; Mandler & Johnson, 1977; Merritt & Liles, 1989; Norbury & Bishop, 2003; Reilly, Losh, Bellugi, & Wulfeck, 2004; Reuterskiöld, Hansson, & Sahlén, 2011; Reuterskiöld Wagner, Sahlén, & Nettelbladt, 1999; Viberg, 2001). In this type of study, the propositions in a child's narrative are compared to propositions in a constructed adult model story. Although closely linked to the narrative content, macrostructure is more than the number of information units expressing story content. Rather, the macrostructure consists of different categories, *macrostructural components*, each with a specific *function* in the narrative. The macrostructure is the core of the narrative, although aspects such as lexicon, syntax and especially referential, temporal and causal linking devices (so-called cohesive devices, Halliday & Hasan, 1976) play a role in narratives, just as they do in all types of discourse.

It should be pointed out that a narrative is not the same as a good (or well-formed) narrative (cf. Stein & Policastro, 1984, p. 151). Most definitions and narrative models are idealized schemata that are more concerned with describing a good narrative (as told by an proficient adult speaker of the language) rather than just anything that fulfills basic narrative criteria (cf. Stein & Policastro, 1984). In the current study, following Berman & Slobin (1994b, p. 44), the term *narrative* is used to refer to productions on narrative tasks. This does not mean that these productions necessary fulfill any specific formal criteria of 'narrativeness'. The term *story* is used for the content that is depicted in the stimulus material, i.e. the picture sequences used to elicit narratives from the children (cf. Section 3.2).

In Table 1.1, the original Labovian model (Labov, 1972; Labov & Waletzky, 1967) is shown. This model, which has been very influential, was developed for the analysis of *personal* narratives, i.e. narratives in which the speaker tells about real events that happened to him/her or another person. In the Labovian model, a central component is the evaluation. The evaluation component contains the reason why the narrative was told (i.e. gives the main point of the narrative). The abstract functions as a foreshadowing summary of the narrative, and the coda creates a link back to the present. The only strictly obligatory part in this model is the complicating action, without which there is not really a narrative (Labov, 1972, p. 370). The Labovian model has later been developed into so-called *high-point analysis* by researchers focusing on children's personal narratives (e.g. Peterson & McCabe, 1983).

Table 1.1. The 'Labovian' narrative model (Labov, 1972, pp. 363, 370).

Component	Description
Abstract	What was this about?
Orientation	Who, when, what, where?
Complicating action	Then what happened?
Evaluation	So what?
Result/resolution	What finally happened?
Coda	Signals that the narrative is finished

Alongside the model for personal narratives, and inspired by traditional fables and fairytales, narrative models were developed for the analysis of *fictional* narratives (e.g. Mandler & Johnson, 1977; Stein & Glenn, 1979). The models for fictional narratives, which were partially based on retell/recall experiments (e.g. Stein & Glenn, 1979), usually focus less (or not at all) on the evaluative component. These models share a minimum formal requirement in addition to the sequential ordering or temporal linking (cf. Hudson & Shapiro, 1991), namely that a narrative consists of a *setting*, and an *episodic system* with at least one episode (e.g. Stein & Glenn, 1979, p. 59), as illustrated in Figure 1.1. The episode is considered to be the core of the narrative. Which components an episode includes varies somewhat between specific models. The setting usually includes time and place of the events as well as introductions of the protagonists, the story characters.

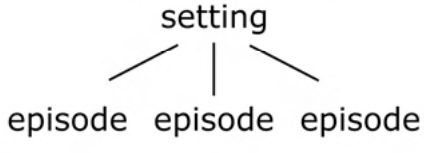

Figure 1.1. Overview of the basic narrative structure for a three-episode narrative.

With the work of Nancy Stein and her colleagues (e.g. Stein & Albro, 1997; Stein & Glenn, 1979; Stein & Policastro, 1984), as well as that of Tom Trabasso and colleagues (e.g. Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992) and Carol Westby (e.g. Westby, 2012), *goals* became central to narrative structure. In the well-known *story grammar* framework,[5] the episodes of a narrative are *goal-based*. This means that each episode is centered around a protagonist's goal: the goal is "the most critical piece of information, for story knowledge is basically organized around the goal of the protagonist" (Stein & Policastro, 1984, p. 118). Table 1.2 shows Stein & Glenn's story grammar model. Alt-

[5] The term story grammar was used because the assumption was that "stories can be described in terms of a hierarchical network of categories and the logical relations that exist between these categories" (Stein & Glenn, 1979, p. 58), i.e. that there are categories and relations for stories in a similar way as constituents and relations in syntax.

hough the goal is central in the model of Stein & Glenn (1979), it forms a part of the internal response category and does not have its own category.[6]

Adapted versions of the Stein & Glenn (1979) story grammar model have been widely used, for example in the analysis of Frog story narratives (e.g. Fiestas & Peña, 2004; Trabasso & Nickels, 1992), where narratives have been scored according to whether or not they include instantiations of the components shown in Table 1.2. The model underlying the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015), shown in Table 1.3, was developed from (goal-based) story grammar models (e.g. Mandler & Johnson, 1977; Stein & Glenn, 1979; Stein & Policastro, 1984), and Westby's (2012) decision-tree model of narrative structure (see below).

Table 1.2. The story grammar model of Stein & Glenn (1979).

Component	My description
Major setting	Introduction of (animate) protagonist(s)
Minor setting	Information that relates to the context of the story (time, place)
Initiating Event	Some type of change occurs for the protagonist
Internal Response	Emotional response to the initiating event (includes the protagonist's goals)
Attempt	The actions carried out by the protagonist
Consequence	Did the protagonist succeed in attaining the goal or not?
Reaction	The protagonist's response to the consequence

Table 1.3. Macrostructural components in the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015).

Component	My description
Setting	Time and place of the events
Internal State as Initiating Event (IST as IE)	What does the character perceive/feel that sets the story events in motion?
Goal (G)	What does the character want?
Attempt (A)	What does the character do (in order to reach the goal)?
Outcome (O)	What is the result? What happens?
Internal State as Reaction (IST as R)	What does the character feel (in response to the outcome)?

As can be seen when comparing Tables 1.2 and 1.3, the MAIN model differs from Stein & Glenn's (1979) model in some respects. The Stein & Glenn

[6] The category *plan sequence*, which consists of the protagonist's *internal plan* as well as the *plan application,* is considered by Stein & Glenn (1979, pp. 60, 64) to be a 'higher order category', i.e. an important building block in the narrative structure. However, in their own analyses, as well as in the description of their model found in Stein & Policastro (1984, p. 118), this category is part of the category *internal response*. For this reason, the plan-component is not included in the model described in Table 1.2.

(1979) internal response and reaction components more or less correspond to the MAIN internal state as initiating event and reaction components, respectively, with the exception that the MAIN internal state component does not include the goal. Instead, the MAIN model has a specific goal component. The MAIN model does not include an initiating event component, and its setting component includes time and place but not the introduction of the characters.[7] These differences mean that results from studies using Stein & Glenn's (1979) story grammar model and those using the MAIN model are not fully comparable.[8] One important advantage of the MAIN model is that it clearly separates goals from internal states, which gives a more detailed picture of the type of inferences speakers draw about characters when telling stories. On the other hand, the fact that story character introductions are not included in the MAIN setting component means that an additional analysis of story character introductions may be needed when the MAIN model is used.

Producing full (or complete) episodes indicates a high level of structural complexity, resulting in a more well-formed narrative (cf. Gagarina et al., 2012, p. 11, 2015; Stein & Glenn, 1979; Westby, 2012). In the story grammar models by e.g. Stein & Glenn (1979), the necessary components in a complete episodic structure are (1) an internal response that gives the motivation *or* the goal of a character *or* an initiating event, (2) an attempt and (3) a consequence (see Stein & Policastro, 1984, p. 119), as shown in Figure 1.2.

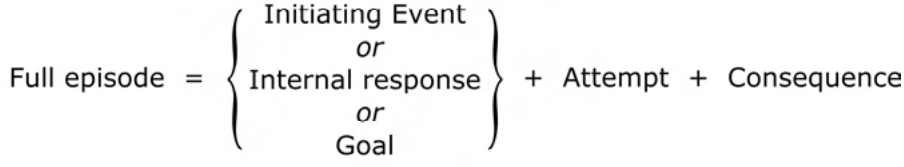

Figure 1.2. The structure of a full (complete) episode in Stein & Glenn's (1979) story grammar model.

Fig. 1.3. The structure of a full (complete) episode in the MAIN story grammar model (Gagarina et al., 2012).

[7] For the different components in a specific MAIN story, see Section 7.3.1.1, Table 7.10. For a more detailed overview of the MAIN model of macrostructure, see Gagarina et al. (2012, pp. 10–13).

[8] Ute Bohnacker's theoretical contributions on this issue are hereby gratefully acknowledged. She not only alerted me to the differences between the Stein & Glenn (1979) model and the MAIN model, but also repeatedly visualized the differences between a full episode in the two models as in Figures 1.2 and 1.3.

The MAIN definition of a full episode is stricter, allowing only a goal as the first component, as shown in Figure 1.3. This means that it may be more difficult to produce a full episode in the MAIN model than in the model of Stein & Glenn (1979). The MAIN model identifies different types of episodic structures (or narrative sequences), with different levels of macrostructural complexity. Based on the decision tree model of Westby (2012, p. 211),[9] episodes can be classified as action/reaction sequences (attempt-outcome sequences), incomplete episodes,[10] which contain a goal but lack either the outcome (goal-attempt sequences) or the attempt (goal-outcome sequences), and complete episodes (goal-attempt-outcome sequence, see Gagarina et al., 2012, pp. 11–12). The different types of sequences are shown in Table 1.4.

Table 1.4. Types of episodic structures (following Westby, 2012).

Type of sequence	**Components**
Action/reaction sequence	Attempt + Outcome (AO)
Incomplete episode	a) Goal + Attempt (GA)
	b) Goal + Outcome (GO)
Complete/full episode	Goal + Attempt + Outcome (GAO)

The types of sequences or types of more or less well-formed episodes can be seen as different developmental stages (cf. Gagarina et al., 2015; Westby, 2012). In the context of picture-based elicited narratives, telling a narrative that contains only action/reaction sequences means that only the actions that are visible in the stimulus material have been included. In order to use incomplete or complete episodes, the speaker needs to draw inferences about the characters' goals from the pictures and verbalize these inferences (cf. Chapter 7). Attributing goals to characters and overtly expressing these is more difficult than telling the events shown in the pictures. This means that using specific types of sequences, i.e. telling a narrative with a specific level of macrostructural complexity, may indicate a certain level of narrative competence (cf. Section 1.2). In the following section, the concept *narrative competence* is described.

1.2 What is narrative competence?

Giving an exact definition of narrative competence is not easy, since it is a complex, multifaceted ability. However, when telling a story, the speaker needs to structure and present the (story) information in a suitable way, taking both the content and the listener's knowledge into account. It has been suggested that narrative competence is the ability both to express the overarching structural organization of the narrative content, the *macrostructure*

[9] See also earlier work by Westby (e.g. Westby, 1984).
[10] In the terminology of Westby (2012), incomplete episodes are called *abbreviated* episodes.

(see Section 1.1), and to use specific linguistic structures at the *microstructural* level (see e.g. Justice et al., 2006; Soodla, 2011). Such specific linguistic structures can for example be the use of a particular (narrative) tense, ritualized story openings (e.g. *once upon a time*), a varied and appropriate vocabulary (for the particular story content), or the use of adequate referring expressions and temporal/causal linking devices, all of which are language-specific. With the exception of the macrostructure, without which there cannot be any properly structured narrative, it is unclear exactly how important each of these aspects is for telling a story, and/or how they are linked to language proficiency in general. Additionally, there are indications that narrative competence is partly universal and partly language- and culture-dependent (Pavlenko, 2006). This means that it is necessary for children to learn to tell a narrative in a way that is appropriate in both the language and the speech community (cultural context) in which it is told.

A number of factors influence children's narrative competence. In fact, it may be more accurate to say that narrative competence consists of a number of (somewhat related) abilities that are essential for being able to tell a good story.

The ability to understand an underlying story schema is a prerequisite for telling a good story. Included in understanding a story schema is the knowledge (or understanding) of goal planning, which is required to be able to include full episodic structures in their narratives (Burris & Brown, 2014; Stein & Glenn, 1979; Trabasso & Rodkin, 1994). As formulated by Trabasso & Rodkin (1994):

> "What is narrated depends upon the narrator being able to interpret the characters and their relations in time and space, to understand how the initiating events impact on the main protagonist and lead to the formation of a goal and goal plan, how the protagonist enacts this plan over time, whether the attempts to achieve the plan fail, how the protagonist reacts to the failure, and finally, how the attempts succeed and end the story. In short, knowledge of goal/plans or planning is required" (Trabasso & Rodkin, 1994, p. 88).

Also central for being able to perform well on picture-based narrative tasks, such as those used in the current study, is the ability to interpret pictures correctly, i.e. to understand story content that is shown visually. This includes to make inferences about the characters' goals and internal states (emotions) (see Burris & Brown, 2014), which is related to Theory of Mind (ToM), i.e. the ability to understand that others have emotions and knowledge that may be different from one's own (cf. e.g. Harris, 1996; Leslie, 1987; Leslie, Friedman, & German, 2004; Tomasello, 2003). Being able to verbalize inferences about story characters and their emotions, is not only central to picture-based narrative tasks, but is an important part of narrative competence in general.

Another important aspect is the ability to link narrative content together in an understandable manner, to structure it in a way that makes it clear to the listener what happened first and what later and who did what (to whom) for what reason(s). This requires the use of cohesive devices (Halliday & Hasan, 1976), such as temporal/causal connectors, and appropriate reference to characters and objects.

In addition to understanding story schemas and using Theory of Mind for making correct inferences, the child also needs to be able to verbalize story content. Naturally, this requires at least some basic level of language proficiency (cf. Viberg, 2001), for example in using suitable, or at least understandable, lexical items and syntactic constructions. Vocabulary has been shown to be especially important for whether narratives are perceived as good (Newman & McGregor, 2006). Vocabulary is also connected to performance on a number of measures of story structure (Heilmann, Miller, Nockerts, & Dunaway, 2010; Uccelli & Páez, 2007). Yet, which level of language proficiency is needed to tell a narrative is unknown and may depend on the complexity of the specific narrative.

All the abilities described above are to some extent related to cognitive maturity (i.e. development of general cognitive abilities), which is in its turn linked to the child's age. In the following section, an overview of the development of narrative competence in children is given.

1.3 Development of narrative competence in the preschool years

This section gives a broad overview of insights from the literature on the overall development of narrative ability in the preschool years, with a focus on ages 3–6/7.[11] Most studies of children's narratives have been carried out on monolingual children, mainly English-speaking children, or on bilinguals speaking a limited number of language combinations, notably English and Spanish.

In the late preschool years, children's narratives develop from loosely attached sequences towards coherent episodic structures (e.g. Berman & Slobin, 1994b; Verhoeven & Strömqvist, 2001). Starting at age 3–4, children's narratives develop from *descriptive sequences*, in which characters and events are described without any explicit temporal or causal linking (e.g. Berman & Slobin, 1994b; Peterson & McCabe, 1983; Westby, 2012), to simple *action sequences*, loosely linked with simple connectives such as *and* or *and then* (e.g. Berman & Slobin, 1994b; Peterson & McCabe, 1983; Westby, 2012), and *reactive sequences* (Stein & Albro, 1997), in which there

[11] Detailed summaries of the literature on vocabulary, character introduction and macrostructure are given in each results chapter (Chapters 5–7).

is causal linking but no mention of characters' goals. Even though children aged 4–6 have generally not been found to produce complete narrative structures (e.g. Peterson & McCabe, 1983), they do use some narrative elements in their stories (see e.g. Berman & Slobin, 1994b; Shapiro & Hudson, 1991; Stein & Glenn, 1979; Trabasso & Nickels, 1992). Around age 5, goals start to appear in the children's narratives, and the proportions of *full episodes*, i.e. goal-attempt-outcome sequences, increase as children grow older (e.g. Trabasso & Nickels, 1992). The following quote summarizes the results from earlier studies of children's narrative competence (for more detailed overviews, see Bohnacker, 2016; Westby, 2012, pp. 202–203):

> "with increasing age and cognitive maturity, there is a gradual move away from descriptive and action sequences and a development toward a more complex episodic organization with causal connections, where the thoughts, feelings, motives, and goals of protagonists, as well as their reactions to successful or failed outcomes, are made explicit for the listener" (Bohnacker, 2016, p. 23).

As children grow older, their narratives thus become more complex as they begin to include more types of macrostructural components, leading to more complete episodic structures. Additionally, they also use more complex linguistic structures and a more suitable lexicon which creates richer narratives. Narrative competence develops beyond the preschool years throughout the school years as well; even narratives of nine-year-olds are not fully adult-like (Bamberg & Damrad-Frye, 1991; Berman & Slobin, 1994c; Pearson, 2002; Schneider et al., 2006; Trabasso & Rodkin, 1994), but this later development falls outside the scope of this thesis and is therefore not described here. The next section describes how narratives can be analyzed in order to try to capture the development taking place from age 4 to 6, the age range of the children in the current study.

1.4 How to study children's narratives: Methodological considerations

When designing a research project to investigate children's narrative competence, there are two main parts of the methodology which both require a number of specific decisions. The first part concerns how to *collect* the narrative data, including which type(s) of narratives to collect, and the second part deals with how to *analyze* the narrative data. This section describes how children's narratives can be studied and simultaneously explains methodological choices made in the current study.

1.4.1 What type of narrative data to collect

In this section, aspects of the data collection process are described, together with the rationale behind the choices made in the current study. Table 1.5 gives an overview of aspects that need to be decided on when collecting narrative data. For each aspect, the choice made in the current study and an alternative are given. For some of the aspects (notably for the type of stimulus), these options are only some out of a range of possibilities.

Table 1.5. Overview of aspects related to collecting narrative data.

Aspects	This study	Alternative
Type of data collection	Elicited	Spontaneous
Type of narrative	Fictional	Personal
Type of stimulus	Picture sequence / booklet	Story stem / single picture
Type of task/mode	Telling	Retelling
Type of listener support	Minimal support	Scaffolding
Type of visual context	Non-shared visual attention	Shared visual attention
Type of language context (for bilinguals)	Monolingual	Bilingual

The first choice to be made concerns the overall *type of data collection*, which could be either spontaneous, if which case the children produce narratives (or not) at a time of their choice, or elicited, which means that the researcher in some way or another makes the children produce narratives at a specific time, e.g. by administering a type of narrative task to the children. Trying to collect data of children's spontaneous narratives is extremely time-consuming. It requires the researcher to follow a group of children for extended periods of time, running the risk of not resulting in any narrative data at all (since the children may not produce any narratives while the researcher is around) while simultaneously collecting a large amount of non-narrative data (i.e. from spontaneous speech). While this type of approach may give valuable insights into what type of narratives children spontaneously produce, it is too difficult to predict what the resulting narratives would look like, what would the total number of narratives would be, how many children would tell a story etc. For these reasons, in the current study only elicited narratives were collected.

Second, it is necessary to decide on the *type of narrative*. Either personal narratives, i.e. narratives about the child's own experiences, or fictional narratives, which consist of made-up events, could be collected. While some researchers advocate the use of personal narratives (e.g. McCabe, 1996; McCabe & Rollins, 1994; Peterson & McCabe, 1983), it is clear that even if a specific topic is given (e.g. a visit to the doctor, a birthday party), children's responses to a request to tell a personal narrative may be very varied.

Some children may even have difficulty telling anything at all. The resulting narratives may thus be difficult to compare. Fictional narratives, especially when the topic and the characters are specified, will be more comparable. Therefore, fictional narratives were used in the current study.

The third aspect to consider is which *type of stimulus* to use. There are various types of stimuli that can be used to elicit narratives, for example story stems, single pictures, or picture sequences. Different stimuli types impose different types of restrictions on the children's production process. A simple stimulus such as a story stem (e.g. "Tell me a story about a boy and his dog") leaves the narrative content open, giving the child the possibility to fully use his/her own imagination in telling the story. However, this may also be difficult for some children, who will not immediately come up with their own story to tell, and then may experience discomfort or even anxiety. Similarly to personal narratives, narratives elicited with a story stem may also be very varied and therefore difficult to compare. Using a single picture depicting one scene from a story as stimulus directs the child's narrative somewhat more, but still leaves parts of the content open. Using a sequence of pictures or a wordless picture book steers the child towards a specific narrative, i.e. the one shown on the pictures, making the resulting narrative data more uniform (cf. Fiestas & Peña, 2004). To maximize comparability, the current study used picture sequences as stimuli.

The fourth choice concerns the *type of task/mode* to use, i.e. the elicitation method. The choice here is whether to have the children retell a story that they have heard (either read live by a researcher or a recorded story over headphones), or to tell the story directly from pictures. There are different opinions in the literature as regards which task is most suitable for studying children's narrative abilities. For example, Schneider, Hayward and Dubé (2006) state that "formulation [telling] appears to provide more information about children's independent storytelling abilities" (Schneider et al., 2006, p. 225). The choice of task may also depend on the purpose of the study. The main reason why telling was chosen in the current study instead of retelling or a combination of the two task types, is that retelling tasks, even those where the child is allowed to look at the pictures both while listening to the story and while retelling, is not only a test of the child's narrative ability. It is also a test of the child's (verbal) memory (cf. Boudreau, 2007; Brookshire & McNeil, 2015, p. 129; Dodwell & Bavin, 2008). As issues related to memory fall outside of the scope of the current study, it was decided not to include a retelling task.

The final three decisions, which all concern the type of context in which the narratives are elicited, are the *type of listener support* given to the child and the *type of visual context* and *language context*. The type of listener support concerns how much help with the narrative task the experimenter should give the child, e.g. in the form of more or less specific prompting. When minimal support is given, as opposed to when the child's narration is scaf-

folded by the experimenter, the child's ability to narrate a series of events independently is measured.[12] Visual context concerns whether there is shared visual attention, i.e. whether the researcher can see the stimuli or not. Having a context of non-shared visual attention means that the child is expected to refrain from using deictic words and/or gestures such as pointing; since the listener cannot see the stimuli, characters and actions need to be introduced explicitly in order for the listener to understand what the story is about. The child can thus not rely on the shared visual context in his/her narration, and this means that a task with non-shared visual attention tests the child's ability to take the listener's perspective into account. Finally, language context, which is a relevant aspect in studies of bilinguals, concerns whether the experimenter speaks to the child in both languages, thus creating a bilingual context, or whether care is taken to create a monolingual context. Creating a monolingual context means that the child is encouraged to use only one language, the language of the testing. In this way, the child's ability to carry out the narrative tasks in one specific language can be assessed.[13]

Taken together, the decisions described here resulted in the following data collection procedure for the current study: fictional narratives were elicited using picture-based stimulus materials following standardized procedures with non-shared visual attention and minimal support given to the children. Care was taken to create a monolingual context for the bilingual children. The specific narrative instruments chosen, the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015) and the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2005) are described in Section 3.2, and the procedure is explained in Section 3.3. The main idea when designing the current study was to create a relatively controlled data collection, since such a procedure will lead to more comparable narratives. In turn, being able to compare narratives between individuals and groups is a clear advantage when studying (bilingual) children's narratives. The reader should keep in mind that the current study elicited narratives in this specific way, using these specific instruments, and that the results may not generalize to all types of narratives elicited with all types of methods, as "[d]ifferences in context, as well as elicitation technique, may impact narrative performance" (Fiestas & Peña, 2004, p. 157).

[12] Note however that this situation is less similar to everyday life communicative situations, in which different types of support are often available from adults.

[13] Investigating how bilinguals narrate in a *bilingual mode* when they use both languages or the language of their choice, would constitute a different type of study, a study that aims to analyze the bilinguals' combined narrative competence. The current study deals with the bilinguals' narrative competence in both languages, but the languages are assessed separately.

1.4.2 What to analyze in narratives

In a project the size of the present study, not all aspects of narratives can be analyzed. This section gives an overview of different aspects of narrative ability that can be analyzed, together with a rationale for the three that were chosen in the current study, vocabulary (Chapter 5), character introduction (Chapter 6) and narrative macrostructure (Chapter 7).[14] The choice of the three aspects was mainly based on them being central to narratives (cf. Newman & McGregor, 2006).

A distinction is often made between two levels of narrative analysis, microstructure and macrostructure, separating aspects on the local (micro) level, which focuses on words, clauses and utterances, from the global (macro) organization of the narrative as a whole (see e.g. Gagarina et al., 2012; Justice et al., 2006; Liles, Duffy, Merritt, & Purcell, 1995). It has been suggested that analyzing both these levels is necessary in order to get a fuller picture of narratives (e.g. Justice et al., 2006; Soodla, 2011). There is consensus in the literature that the global level is the same as the overall structure of the narrative, i.e. the macrostructure (as described in the narrative models in Section 1.1). Being able to express essential macrostructural components can be seen as the core of narrative competence (cf. Section 1.2) and it develops extensively throughout the ages covered by the current study (cf. Sections 1.3 and 7.1). Therefore, macrostructure is a central aspect to analyze in a study of narratives.

In contrast to the global level, the aspects included in the local level vary. The local level usually contains a wide variety of different linguistic aspects. These aspects range from how clauses and information are linked, i.e. temporal, causal and referential cohesion, via the use of internal or mental state terms, words (e.g. adjectives and verbs) used to describe what Bruner (1986) termed the *landscape of consciousness* (what story characters think and feel) to a number of more general linguistic aspects (for an overview, see e.g. Soodla, 2011, pp. 11–12). These more general aspects include measures of productivity, grammaticality, syntactic complexity, and lexical diversity.[15]

Narratives do not only contain descriptions of events, but can also include evaluative statements (cf. Bamberg & Damrad-Frye, 1991; Labov & Waletzky, 1967; Peterson & McCabe, 1983), including statements about the internal (or mental) states of characters. Internal state terms, and other types of evaluative statements, help create a richer narrative, because not only actions are described but also characters' internal responses to these actions, e.g. in the form of emotions. In the MAIN model (see Section 1.1), the internal states thought to be most central (those as initiating event and as reac-

[14] How each of the aspects analyzed in the current study were coded and which specific analyses were carried out is described in each results chapter (Chapters 5–8).

[15] Naturally, narrative data can also be used to analyze the occurrence of specific syntactic and/or lexical phenomena, e.g. relative clauses or the use of specialized vocabulary.

tion) are included as macrostructural components. In the current study, it was therefore not deemed necessary to conduct a separate analysis of internal state terms.

The cohesion of narratives is often investigated by analyzing the use of cohesive devices. Cohesive devices are markers of e.g. temporality/causality and reference that link a text/narrative together (Halliday & Hasan, 1976), such as connectives and referring expressions. Contrary to other models of macrostructure (cf. Section 1.1), introduction of story characters is not scored as part of the MAIN setting component. However, properly establishing characters in the discourse is important for narrative cohesion. Without clear introductions of characters, the rest of the narrative will be difficult to follow. For this reason, a separate analysis of character introduction was included in the current study.

In a study of ratings of children's narratives, Newman & McGregor (2006) found that, in addition to story grammar (macrostructure), both teacher and layperson raters paid more attention to vocabulary than to syntax, for example, when rating children's narratives. Choosing lexical items that are suitable for the narrative content, i.e. words that accurately describe the different events, is important for telling a good story. It thus seems to be the case that not only the inclusion of story grammar elements, i.e. macrostructural components, but also vocabulary is central to what constitutes a good narrative. Additionally, some earlier studies have shown a link between narrative vocabulary and story structure (Heilmann et al., 2010; Uccelli & Páez, 2007). Therefore, in the current study narrative vocabulary was investigated. A general measure of vocabulary (lexical knowledge), such as scores on a vocabulary test, can function as an indication of the child's general language proficiency, which forms a prerequisite for being able to tell a narrative. For this reason, the current study included an analysis of the children's scores on lexical tasks (see Section 3.2.1) in addition to narrative vocabulary.

1.5 How to investigate narrative competence in bilinguals

A number of additional aspects need to be taken into account when studying narrative competence in bilinguals compared with studying monolinguals. In addition to the choice of language context, as described above (Section 1.4.1), three aspects are especially important. First and foremost, the development of a bilingual child's languages may differ from that of monolingual children. Bilinguals often show different levels of knowledge in the two languages, there is cross-linguistic influence (transfer) between the languages, and a larger individual variation (Kohnert, 2010). Therefore, it is not

advisable to describe the linguistic ability or narrative development of a bilingual child using monolingual norms and data only from only one language; data from both the child's languages are needed. Although a number of studies have found similar results for narrative macrostructure in the two languages of bilingual children (e.g. Bohnacker, 2016; Fiestas & Peña, 2004; Pearson, 2002), this does not mean that investigating just one language is enough to gain a full picture of bilinguals' narrative competence. The first important point is therefore to investigate narratives in bilinguals' both languages.

Second, since two languages are to be investigated and results from the two languages are to be compared, it is important to use equivalent narrative stimulus materials (stories) in the two languages. Some earlier studies used the same story in both languages or used non-comparable stories in the two languages. Children may find it boring to tell the same story twice, and/or there may also be practice/learning effects from the first to the second testing (cf. Pavlenko, 2008). If non-comparable stories are used, results for the two languages are not equivalent. Therefore, different but comparable stories should be used in the two languages. The stories need to be counterbalanced across languages. To put it simply, this means that half of the children tell story 1 in language A and story 2 in Language B and the other half tell story 2 in language A and story 1 in language B.[16]

Third, it is crucial to balance the order of the languages. This means that half of the children are tested in one language first and the other half in the other language first. In this way, if there is any training effect from the first to the second testing, it will not only affect performance in one language. If the languages were not balanced, any training effect would make results for the language that was tested second higher than they would otherwise have been.

In the current study, all these three aspects have been taken into account. Both languages of the bilinguals were investigated, using pairs of stimulus materials that have been carefully constructed to be as similar as possible in terms of story structure as well as number and type of characters. Half of the bilinguals were tested in the majority language Swedish first and the other half in the respective minority language, German or Turkish, first.

[16] For an overview of the counterbalancing system used in the current study, see Section 3.3.1.

2 Aims and research questions

The current study has two main aims. The first aim is to gain insights into the development of oral narrative competence from age 4 to 6 in monolingual Swedish children and in both languages of Swedish-German bilinguals and Swedish-Turkish bilinguals. The study investigates three different narrative aspects: vocabulary, character introduction and (comprehension and production of) narrative macrostructure. Focus is on preliterate development of oral narrative competence and thus the study mainly includes children who cannot read or write. Different age groups (four-, five- and six-year olds) are compared. A comparison of these years is relevant, as during this age range children's language proficiency and cognition develop. Since few large-scale studies have been carried out on bilingual children's narrative competence and development in both languages, and none in a Swedish context, the aim is not to test one particular hypothesis or theory, but rather to explore children's narratives from different angles.

In order to assess children's narrative competence, it is necessary to have adequate tasks and procedures. This requires more knowledge about task effects. The second aim is therefore to investigate task effects. To this end, performance on different narrative tasks will be compared, both in comprehension and production.

Four research questions are asked in the current study. The first three fall under the first aim, namely to investigate the children's narrative competence. The last research question belongs to the second aim, which concerns task effects. Each of the research questions is first explored separately for vocabulary, character introduction and macrostructure, after which the results are related to each other in the general discussion.

- **RQ1**: How and when do the different aspects of narrative competence (vocabulary, character introduction, macrostructure) develop between age 4 and 6?
- **RQ2**: To what extent is this development similar for mono- and bilingual children and for different bilingual groups?
- **RQ3**: Are there differences between the bilingual children's two languages?
- **RQ4**: Are there differences in the children's performance depending on the narrative task?

3 Method

This chapter deals with the method of the present study. It describes the participants of the study (Section 3.1), the stimulus materials used for eliciting the data (Section 3.2), the data collection procedure (Section 3.3) and transcription (Section 3.4). Descriptions of coding and analysis for each of the narrative aspects studied are found in each results chapter (Chapter 5 Vocabulary, Chapter 6 Character introduction, Chapter 7 Macrostructure, Chapter 8 Linking character introduction and vocabulary to macrostructure).

3.1 Participants

In this section, the participants of the study are described, the monolingual participants (Section 3.1.1) and the participants in the two bilingual groups, the Swedish-German and Swedish-Turkish children (Section 3.1.2). The parents of all participants signed a consent form and filled in a questionnaire. In addition to giving information about the social and linguistic backgrounds of the children, the aim of the questionnaire was to ensure that the children fulfilled the criteria for participation in the study, i.e. that they spoke the required languages and had typical language development. The focus of the current study is narrative competence in (bilingual) children who are typically-developing. Therefore, children who had received a diagnosis of language impairment (LI)[17] or were suspected to have such language problems were excluded from the study.

The questionnaire given to the parents of the monolingual children was brief, with questions about languages spoken by the child and in the home (to ensure that these children were indeed monolingual), about education, occupation and languages of both parents, as well as a few general questions about the child's language development. The bilingual questionnaire, which could be filled in by the parents in the language of their choice, Swedish or German/Turkish, was much more detailed. In addition to similar questions as

[17] Children with specific language impairment (SLI) are "children who show a significant limitation in language ability, yet the factors usually accompanying language learning problems – such as hearing impairment, low non-verbal intelligence test scores, and neurological damage – are not evident" (Leonard, 2000, p. 3). In Sweden, there is a diagnosis called *generell språkstörning* 'general language impairment' and various subtypes, but no diagnosis of SLI. For this reason, the term LI is used here.

those asked for the monolinguals, it also contained questions about the child's exposure to both languages (and other languages), and the child's proficiency in the languages.

Each child was given a unique code, consisting of five letters for the language group (MoSwe = Swedish monolinguals; BiGer = Swedish-German bilinguals; BiTur = Swedish-Turkish bilinguals), one digit for the age group (4, 5, or 6), and a two-digit number identifying the individual child. To give some examples, MoSwe4-19 is a monolingual four-year-old and BiGer6-04 is a Swedish-German six-year-old. A code was given to all children that were tested.[18] These codes are used whenever reference is made to specific children.

3.1.1 Monolinguals

The monolingual data was collected March – December 2014. The 72 monolingual children were recruited from seven preschools and three schools in two larger cities in central Sweden. They were recruited with the help of preschool and school staff who contacted parents and collected consent forms. All monolingual participants fulfilled the same basic criteria for participation in the study: no knowledge of any other language than Swedish and no diagnosed language impairment. All four- and five-year-olds attended preschool and all six-year-olds attended a preparatory school year before Grade 1 called *förskoleklass* 'preschool class'. Table 3.1 gives an overview of the monolingual participants.

Table 3.1. Overview of the monolingual participants.

Monolinguals	**4-year-olds**	**5-year-olds**	**6-year-olds**	**Total**
N	24	24	24	72
Girls/boys	12/12	12/12	13/11	37/35
Mean age	4;5	5;6	6;5	5;5
Age range	4;0–4;10	5;0–5;11	6;0–6;10	4;0–6;10

None of the children had been diagnosed with language impairment and there was no indication that any of them had a neuropsychiatric disorder, e.g. ADHD or autism-spectrum disorder or any other type of disability. A few parents reported concern about their child's language development. In all cases, this was due to the child having had unclear speech or other problems with pronunciation. Five children were reported to have some language problems in the family. In three cases, this was a parent or sibling, with pronunciation problems and in two cases a family member had dyslexia. Three children had had some problem related to hearing in their early childhood. In

[18] The same system for assigning codes to participants was also used in the BiLI-TAS project (Bohnacker, 2013), the research project at Uppsala University within which the Swedish-Turkish data was collected (PI: Ute Bohnacker).

all cases, this was due to frequent ear infections. The results of these children did not diverge from other children in their age group and they were therefore included in the study. All 72 children were reported to have normal hearing at the time of the testing.

Five children (6.9%) had one parent with another first language than Swedish. The languages were Slovene, Polish, Bosnian, Danish and Finnish. These languages were not spoken in the home and none of the parents reported that the child spoke another language than Swedish. All the tested children can therefore be included in a sample of monolingual Swedish-speaking children.

The children were divided into two groups based on the socio-economic status (SES) or their parents, *high* and *low* SES. Parental education was used as a proxy for SES. Parents' answers to an open-ended question about their education were coded according to the United Nations ISCED 2011 classification (UNESCO Institute for Statistics, 2012). This classification has 9 levels ranging from 0 (= 3 years of early childhood education) to 8 (= PhD). Parents who were doctoral students were included in level 8 and master students were included in level 7 (= MA/MSc). On the basis of answers from both parents, average parental education was calculated. Average parental ISCD levels 0-3 (education up to and including completed secondary education) were combined into the category low SES and levels 4-8 (any tertiary education up to and including PhD) into high SES. Children whose average parental education was 3.5 (which means that, on average, one parent had completed secondary education and one parent had some further education, but not at university level) were placed in the low-SES group. When data was only available for one parent, SES was assigned based on the educational level of that one parent. For a justification of this way of calculating SES, see Bohnacker, Lindgren & Öztekin (2016, pp. 27–28). The mean parental education for the monolingual participants was 5.6 (SD = 1.6), which means that, on average, the children had one parent with a BA and one parent with some tertiary education but not at university level. Fourteen children (19.4%) belonged to the low-SES group, whereas the rest of the children (59 children; 81.9%) had a high-SES background. Generally, in Sweden, 47% of adults aged 25–44 (the age group to which the parents of the children in the current study belong) have some form of tertiary education (Statistics Sweden, 2017a, p. 7). Thus, a considerably larger proportion of the monolingual participants of the current study have a high-SES background than in the population at large.[19]

[19] There are indications that SES influences some aspects of language development, notably vocabulary, in monolinguals (cf. Hart & Risley, 1995).

3.1.2 Bilinguals

3.1.2.1 General Information

The bilingual data was collected between November 2014 and October 2016. All bilinguals included in the study fulfilled the same basic criteria, namely that they were growing up in Sweden speaking Swedish and either Turkish or German, that they were able to complete the testings in both languages, and that they did not have any diagnosed language impairment. Forty-six Swedish-German bilinguals and 48 Swedish-Turkish bilinguals aged 4;0–6;11 participated. An additional eight Swedish-German bilinguals (six four-year-olds, two six-year-olds) were tested, but could not be included in the study as they were either unwilling or unable to complete the testing in German. In the German testing, these children spoke too limited German to complete the tasks although they did have a substantial passive knowledge of the language (i.e. they seemed to understand most of the instructions given by the experimenter).

The Swedish-Turkish bilinguals were a subsample of the participants in the BiLI-TAS project (Bohnacker, 2013). Sixteen participants in each age group, i.e. a total of 48 participants, were selected from the 74 Swedish-Turkish BiLI-TAS participants who were within the age range of the current study (4;0–6;11). The reason for selecting 48 children instead of including all 74 was to have a comparable number of participants in both bilingual groups. Children were selected based on which narratives they had told, so as to conform to the current study's counterbalancing system for the narrative tasks (see Section 3.3.1). Only children who completed the tasks in both Swedish and Turkish were selected.

The bilingual participants were recruited using various channels, such as by contacting mother-tongue teachers, who in turn contacted parents of children in the right age range, by contacting preschools (especially for the Swedish-Turkish bilinguals), and by using personal contacts and social media. The bilingual participants were recruited from the larger Stockholm area and from another larger city in the same region. Two Swedish-German children and four Swedish-Turkish children were siblings. Most six-year-olds attended *förskoleklass* and most children in the younger groups attended preschool. In the Swedish-German group, ten children attended a bilingual Swedish-German preschool and four children attended *förskoleklass* at a Swedish-German bilingual school. All other children attended Swedish-medium preschools and schools.

A general overview of the bilingual participants is given in Table 3.2. A one-way ANOVA showed no significant age differences between the monolinguals, the Swedish-German bilinguals and the Swedish-Turkish bilinguals ($F(2, 164) = 0.274$, $p = .761$).

Table 3.2. Overview of the bilingual participants.

Swedish-German bilinguals	**4-year-olds**	**5-year-olds**	**6-year-olds**	**Total**
N	14	16	16	46
Girls/boys	9/5	12/4	10/6	31/15
Mean age	4;6	5;5	6;7	5;7
Age range	4;0–4;11	5;0–5;11	6;1–6;11	4;0–6;11
Swedish-Turkish bilinguals	**4-year-olds**	**5-year-olds**	**6-year-olds**	**Total**
N	16	16	16	48
Girls/boys	8/8	10/6	9/7	27/21
Mean age	4;6	5;7	6;5	5;6
Age range	4;0–4;11	5;0–5;11	6;0–6;11	4;0–6;11

All participants were active bilinguals, as illustrated by their ability to complete the tasks in both languages. A number of children spoke another language in addition to the two languages investigated here. One Swedish-Turkish bilingual also spoke Dimili (Zaza). In two families, Kurdish (Kurmanji) was spoken to a relatively large extent, so these children were trilinguals. In three other families, the parents also had Kurdish as first language (L1) and spoke a mix of Turkish and Kurdish to each other. One can therefore assume that these three children had at least some passive knowledge of Kurdish. Eight of the Swedish-German children were trilinguals, with two children speaking Chinese and five children speaking English as their third language. The last Swedish-German trilingual spoke some Estonian. Trilinguals were not excluded, as the study seeks to gain an understanding of narrative ability in children speaking Swedish and German or Turkish and, as the participants of the current study show, there are trilinguals in both these groups. Due to the limited availability of children with these two language combinations, and the wish to have relatively large numbers, it was also for practical reasons not possible to exclude trilinguals. The trilinguals fulfilled the same criteria as the bilinguals, i.e. they were able to complete the tasks in Swedish and German/Turkish. No systematic differences between the bilingual and the trilingual participants could be found for any of the studied measures; some trilinguals performed well and some less well in one or both of the studied languages. It was therefore decided not to analyze the data from the trilinguals separately.

Almost all Swedish-Turkish children (94%, 45 children) and the large majority of the Swedish-German children (74%, 34 children) had been living in Sweden since birth. Three Swedish-Turkish bilinguals were born in Turkey. Five Swedish-German bilinguals were born in Germany and one in Switzerland, and five in other countries. Additionally, one Swedish-German five-year-old had been adopted from South Korea at the age of seven months. All children except two Swedish-Turkish children and two Swedish-German children had been living in Sweden for at least two years. As chil-

dren born in other countries form a natural part of most bilingual groups in Sweden, these children were not excluded, as long as they fulfilled the primary criterion of being able to complete the tasks in both languages. The initial goal of the study was to only include children who had been living in Sweden for at least two years, but when a few children with shorter stays were recruited, and they were able to speak enough Swedish to complete the tasks, they were also included in the study.

3.1.2.2 Parental concern about language development

Information about (potential) problems related to language was collected from the parents to ensure that all children included in the study had typical language development and thus fulfilled the criteria for inclusion in the study. None of the bilingual children had been diagnosed with language impairment or any type of neuropsychiatric disorder, e.g. ADHD or autism-spectrum disorder.

An indicator of the child's language development is the presence or absence of parental concerns (Salameh, Nettelbladt, & Gullberg, 2002). The parents of seven Swedish-German and eleven Swedish-Turkish bilinguals had experienced concern about their child's language development. The majority of worries concerned just one language[20] and none of the concerns seem to give reason for excluding children from the study.

A number of children in each bilingual group (Swedish-German: five children; Swedish-Turkish: six children) had been in contact with a speech-language pathologist (SLP).[21] However, the overall impression of these children both from the parental reports and from test scores gives no reason to suspect language impairment in any of these children; they can thus be included in the study.

[20] Three families (one Swedish-German, two Swedish-Turkish) worried about the child's pronunciation. The parents of BiTur6-21, who had a hearing aid (see below), had been worried because of the child's hearing problems. Three families were concerned because the child had a generally late language development. One family wrote that they had started to worry after age two. In the case of five children, parents were concerned about the minority language, and the fact that the child did not speak it very well or spoke mainly Swedish. Three Swedish-Turkish families were worried about their child's Swedish, either because his/her development in Swedish had been late, or because the child was shy/insecure when speaking Swedish, and one additional family in this group had been concerned about their child not being able to manage at school, presumably due to late language development in Swedish. One Swedish-German family was worried because there were a lot of languages in the family.

[21] In the case of BiTur6-21, this was linked to the child being fitted with a hearing aid. Three Swedish-German children and one Swedish-Turkish child had visited an SLP due to pronunciation difficulties, or in one case stuttering. Four children, one Swedish-German and three Swedish-Turkish, were 'late talkers' and had been in contact with SLPs because of this. The parents of the last Swedish-Turkish child stated that they had seen an SLP to receive extra help with language, but it was not clear for what reason. Finally, one Swedish-German child had seen an SLP because he was often angry at preschool, as a result of having difficulties expressing himself, presumably in Swedish (the staff reported that the child did not know any Swedish when starting preschool at age 1;1).

Language impairment is to a large extent hereditary and a higher rate of language-related problems has been reported in the families children with language impairment than in controls (cf. Kalnak, Peyrard-Janvid, Sahlén, & Forssberg, 2012), Therefore, it is necessary to say a few words about language problems in the families of the bilingual participants. Six Swedish-German and two Swedish-Turkish families reported language problems in the family. In the case of seven of these children, the reported language problems were relatively minor and concerned more distant relatives, e.g. a cousin/aunt with dyslexia. There were no indications that any of these children had delayed language development. Therefore, they can be included in the study. The last child, a Swedish-Turkish six-year-old (BiTur6-09) had an older brother with what the family called *språkstörningar* 'language impairments'. In order to make sure that BiTur6-09 did not have the same problems, the parents' answers to other questions as well as the child's vocabulary scores in both languages were carefully checked. The parents stated that the child had normal development in both languages, and that they had not experienced any anxiety regarding her development. Her vocabulary scores were among the highest in the Swedish-Turkish group in both languages (and within the monolingual range for Swedish). There is therefore no indication that this child might have language impairment, and she can be included in the study.

Two children in each bilingual group had had some problems related to hearing. In the case of two children (BiTur6-02, BiGer4-18), they had had frequent ear infections earlier in childhood. The parents of BiTur6-02 reported that the child had had some problems with pronunciation. The parents of BiGer4-18 reported that the child had received grommets, but there had been no contact with an SLP. This child did have somewhat unclear speech, especially in German. Additionally, one Swedish-German child (BiGer4-14) was reported to have somewhat impaired hearing, and one Swedish-Turkish child (BiTur6-21) had a hearing aid. The parents of BiGer4-14 did not express anxiety about the child's language development, had not had any contact with an SLP and stated that the child's language(s) did not differ from other children's. The child BiTur6-21, whose parents showed some concern had vocabulary scores that were close to the mean scores for Swedish-Turkish six-year-olds in both languages. There are thus no reasons for treating the data from these children differently than those of the other participants.

3.1.2.3 Age of onset for Swedish and rated language proficiency in both languages

All bilingual children in this study had been exposed to the minority language Turkish or German continuously from birth, but the age at which they started receiving regular input in Swedish varied, as shown in Table 3.3. There are important differences between the two bilingual groups. The majority of the Swedish-German children (69.9%, 32 children) had regular in-

put in Swedish from birth or before age 1;0, and only four children after age 2;0. The Swedish-Turkish children were relatively evenly distributed between age of onset at birth or until 1;0 (25%), between age 1;0 and 2;0 (33.3%) and between 2;0 and 3;0 (25%). Eight Swedish-Turkish children had an age of onset after 3;0. This indicates that the Swedish-German bilinguals have received Swedish input for a longer time than the Swedish-Turkish bilinguals, which is partly linked to many of the Swedish-German children having one Swedish-L1 parent, as shown below (Section 3.1.2.4).

Table 3.3. Age of onset of regular exposure to Swedish (%, number of children in parentheses).

	Swedish-Turkish bilinguals	**Swedish-German bilinguals**
Birth to age 1;0	25.0% (12)	69.9% (32)
Age 1;1 to 2;0	33.3% (16)	21.7% (10)
Age 2;1 to 3;0	25.0% (12)	4.3% (2)
Age 3;1 to 4;0	12.5% (6)	0.0% (0)
Age 4;1 – 5;0	2.1% (1)	2.2% (1)
Age 5;1 – 6;0	2.1% (1)	2.2% (1)
Total	**100% (48)**	**100% (46)**

In Table 3.4, the children's ability to speak both languages, according to parental report, is shown.

Table 3.4. Child's expressive language proficiency in the two languages as rated by parents (%, number of children in parentheses).

	Swedish-Turkish bilinguals		**Swedish-German bilinguals**	
	Turkish	**Swedish**	**German**	**Swedish**
very good	58.3 (28)	29.2% (14)	32.6% (15)	71.7% (33)
good	16.7 (8)	47.9% (23)	37.0% (17)	21.7% (10)
'so-so'	16.7% (8)	6.3% (3)	23.9% (11)	4.3% (2)
bad	8.3% (4)	14.6% (7)	4.3% (2)	2.2% (1)
very bad	0.0% (0)	2.1% (1)	2.2% (1)	0.0% (0)
Total	**100% (48)**	**100% (48)**	**100% (46)**	**100% (46)**

There are relatively large differences between the groups. Most Swedish-German children were rated to be more proficient in Swedish than in German, whereas in the Swedish-Turkish group, the Turkish speaking proficiency was rated higher. In both groups, and for both languages, most children were thought by their parents to be 'good' or 'very good' at speaking the language. Eleven Swedish-German children and eight Swedish-Turkish children were rated to only have 'so-so' proficiency in the minority language. In the Swedish-Turkish group, eight children were rated as speaking Swedish badly or very badly; the corresponding figure for Turkish was four children. Only one Swedish-German child for Swedish and three for German were

thought by their parent to have bad or very bad spoken proficiency. The children who were rated as having low language proficiency in one language were still able to complete the testing in that language, and were therefore not excluded from the study.

When the child's proficiencies in the two languages are combined, there are similar numbers of children who are reported to have *very good* proficiency in both languages: nine children (18.8%) in the Swedish-Turkish group and eleven children (23.8%) in the Swedish-German group. Eight Swedish-Turkish and two Swedish-German children were reported to have relatively low proficiency in both languages. The two Swedish-German children were both four-year-olds. In the Swedish-Turkish group, it was more common for five- and six-year-olds (three and four children, respectively) to be reported to have low proficiency in both languages, than among the four-year-olds (one child). The majority of the children, 31 Swedish-Turkish children (64.5%) and 33 Swedish-German children (71.8%) fell in between these two extremes, having one stronger language or two languages that were rated as 'good'.

Parents were also asked if one of the child's languages was stronger at the time of testing. The results, shown in Table 3.5, point to important differences between the groups. In fact, the pattern found for the Swedish-Turkish group was the opposite of the Swedish-German group. The number of children who were rated to have equal proficiency in both languages was similar in the two groups, but the distribution between those having Swedish and those having German/Turkish as the stronger language were radically different in the two groups.

Table 3.5. The child's strongest language at the time of testing according to parental report (%, number of children in parentheses).

	Swedish-Turkish bilinguals	**Swedish-German bilinguals**
Both languages equal	39.6% (19)	32.6% (15)
German/Turkish	45.8% (22)	13.0% (6)
Swedish	14.6% (7)	54.3% (25)
Total	**100% (48)**	**100% (46)**

The parents of 20 Swedish-Turkish and 24 Swedish-German children stated that the child had a preference for one of the languages. Whereas only two Swedish-German children preferred German, and 22 preferred Swedish, in the Swedish-Turkish group, a similar number of children preferred each language (Swedish: 9 children, Turkish: 11 children). The remaining children (27 Swedish-Turkish and 21 Swedish-German bilinguals) had no clear preference for any of the languages, according to their parents.[22]

[22] Note that data was missing from one Swedish-Turkish family.

To summarize, in both groups, there are children with higher proficiency in one of the languages, and children who prefer one of the languages. In both groups, most children had 'good' or 'very good' proficiency in both languages. The majority of children were exposed to Swedish before age 3 (the minority language Turkish/German was always spoken in the family). However, there were also differences between the two bilingual groups, with the Swedish-German children generally being exposed to Swedish earlier and having a higher proficiency rating in Swedish than the Swedish-Turkish children. It was more common in the Swedish-Turkish group for the children to have Turkish as their stronger language, but in both groups, a relatively high number of children were rated by their parents to have equal proficiency in both languages, and to have no clear preference for one language.

3.1.2.4 Social and linguistic backgrounds of the parents

There were large differences in SES between the two bilingual groups.[23] The ISCED 2011 mean of the average parental education of the Swedish-Turkish bilinguals was 3.5 (SD = 1.6, range = 1–7.5); the average Swedish-Turkish child had parents who had completed secondary school, with one parent having some additional education. The range of educational levels was very wide in this group, covering almost the entire scale, with some children having parents with only six years of primary schooling, and others parents with master degrees and in one case a PhD. Two-thirds of the Swedish-Turkish bilinguals (65%, 31 children) belonged to the low-SES group and one third (35%, 17 children) had high-SES backgrounds.

By contrast, all Swedish-German children came from high-SES backgrounds. The average Swedish-German child had parents who had both completed at least a bachelors' and in most cases a masters' degree (mean average level of parental education 6.9, SD = 0.9, range = 4–8). All Swedish-German parents had some post-secondary education, and 27 parents had a PhD or were doctoral students. The large difference in SES between the two bilingual groups, as well as how they differ from the monolinguals, needs to be kept in mind, as SES has been shown to influence aspects of language, such as vocabulary, in both mono- and bilingual children (Buac, Gross, & Kaushanskaya, 2014; Calvo & Bialystok, 2014; Hart & Risley, 1995). According to Statistics Sweden (2017a, pp. 32–33), 19% and 60% of the people living in Sweden aged 25–64 born in Turkey and Germany, respectively, had some post-secondary education. These figures are likely higher for the age group the parents of the children in the current study belong to (most are probably in the age range 30–45). Also, a substantial part of the children in the current study had (at least) one parent born in Sweden. However, it still seems likely that the children participating in the current study have a higher SES than the average child with the same linguistic

[23] For information about how SES was calculated, see Section 3.1.1.

background. This is probably an effect of the sampling procedure; parents with higher education may be more likely to sign up their child for such a study.

Another important aspect for understanding the bilingual participants is the linguistic backgrounds of their parents. This includes information about where the parents were born and grew up as well as the length of stay in Sweden for parents not born there. In Table 3.6, the countries were the children's parents were born and grew up are shown.

Table 3.6. Parents' countries of birth and growing up (%, number of parents in parentheses).

	Swedish-Turkish bilinguals (N=48)	**Swedish-German bilinguals (N=46)**
Born and grown up in Sweden	22.9% (22)	27.2% (25)
Born and grown up in country where ML is spoken	66.7% (64)	60.8% (56)
Born in country where ML is spoken, grown up in Sweden	5.2% (5)	0.0% (0)
Born in Sweden, grown up in country where ML is spoken	0.0% (0)	1.1% (1)
Other	2.1% (2)	10.9% (10)
Missing information	3.1% (3)	0.0% (0)
Total	**100% (96)**	**100% (92)**

Note. ML = minority language (German or Turkish).

Most parents were born and grew up in the same country. For the majority of the parents, this was a country where German/Turkish was spoken.[24] A substantial proportion of the parents was born and grew up in Sweden. Five Swedish-Turkish parents were born in Turkey, but had spent most of their childhood in Sweden. One Swedish-German parent was born in Sweden, but grew up in Germany. Additionally, a number of parents belong to the category other, having spent their childhood in different countries.

In the Swedish-Turkish group, the average length of stay in Sweden of parents not born in Sweden was 13.3 years (SD = 8.6), and the length of stay varied from 0.6 years to 41 years. The corresponding figure for the Swedish-German group was 9.4 years (SD = 5.0), with stays from one year to 25 years.

In Table 3.7, the first languages (L1s) of the parents are shown. There are important differences between the groups. In the Swedish-Turkish group, it was rare to have Swedish as L1, but this was relatively common among the Swedish-German parents. Only four Swedish-Turkish children had one parent with L1 Swedish, compared with 25 Swedish-German children. Additionally, three parents in the Swedish-German group were Swedish-German

[24] Two Swedish-German parents were born and grew up in Switzerland, one in Austria, and the other 53 parents were born and grew up in Germany.

bilinguals themselves. In both groups, a number of parents had another language than Swedish or German/Turkish as their L1.[25]

Table 3.7. Parents' L1s (%, number of parents in parentheses).

	Swedish-Turkish bilinguals (N=48)	**Swedish-German bilinguals (N=46)**
SWE	4.2% (4)	27.2% (25)
Bilingual SWE & ML	0.0% (0)	3.3% (3)
ML	78.1% (75)	63.0% (58)
Other	13.5% (13)	6.5% (6)
Missing information	4.2% (4)	0.0% (0)
Total	**100% (96)**	**100% (92)**

Note. SWE = Swedish, ML = minority language (German or Turkish).

3.1.2.5 Language use in the home

Three important aspects of language use in the home can be distinguished. The first is language use between parents, i.e. the language(s) parents use to communicate with each other. The second aspect is the language(s) parents use to speak to the child. Together, these two aspects constitute the child's *parental language input*. The third aspect, namely the child's own language use with the parent, is also central, as it contributes to the family language environment (cf. the input–proficiency–use cycle, Pearson, 2007).

In 33 Swedish-Turkish families (73.3%) and 20 Swedish-German families (43.5%) the parents reported speaking to each other exclusively in the minority language Turkish/German. In five further Swedish-German families (10.9%), the language between the parents was mostly German. In a relatively small number of families, five Swedish-Turkish (11.1%) and three Swedish-German families (6.5%), a mix of Swedish and Turkish/German was used. In some families, parents spoke to each other in a third language: either exclusively (one Swedish-Turkish family, Kurdish; three Swedish-German families, English) or a third language together with the minority language Turkish/German (five Swedish-Turkish families: one family Dimili (Zaza), four families Kurdish; two Swedish-German families: English). In three Swedish-German families (6.5%), Swedish was mostly used between the parents. In the remaining ten Swedish-German families (21.7%), the parents spoke to each other only in Swedish. This was only the case in one Swedish-Turkish family (2.2%). Information was missing from three Swedish-Turkish families, but it seems likely that these families did not differ from

[25] Ten of the parents in the Swedish-Turkish group had Kurdish as their L1, and one more was a Kurdish-Turkish bilingual. These ten parents formed five couples, so that five of the Swedish-Turkish bilinguals grew up in a home where both parents spoke Kurdish, and where no parent had Turkish as their first language. In all these families, Turkish was spoken more frequently than Kurdish (according to the parental questionnaires). The last two parents in this group had Dimili (Zaza) as L1. In the Swedish-German group, three parents were bilinguals, having grown up with German/English, German/Italian and Swedish/Estonian. Two parents had L1 Chinese and one was a native speaker of Polish.

the general pattern found in this group, i.e. parents using primarily Turkish with each other.

Table 3.8 shows the parents' reported language use with the child.[26]

Table 3.8. Parents' reported language use with child (%, number of children in parentheses).

	Swedish-Turkish bilinguals (N=48)	**Swedish-German bilinguals (N=46)**
1 almost only SWE, 1 mostly SWE	0.0% (0)	2.2% (1)
1 mostly SWE, 1 50/50	4.3% (2)	2.2% (1)
OPOL (1 mostly/almost only ML, 1 mostly/almost only SWE)	10.6% (5)	50.0% (23)
Both 50/50	6.4% (3)	2.2% (1)
1 mostly/almost only ML, 1 50/50	8.5% (4)	2.2% (1)
Both almost only/mostly ML	68.1% (32)	34.7% (16)
Other[27]	2.1% (1)	6.5% (3)
Total	**100% (47)**	**100% (46)**

Note. SWE = Swedish, ML = minority language (German or Turkish), OPOL = one parent, one language.

Table 3.8 shows large differences between the two bilingual groups. Whereas in more than two thirds of the Swedish-Turkish families, both parents reported speaking almost only or mostly in the minority language to the child, this was the case in only around one-third of the Swedish-German families. Instead, it was most common for one parent to speak German and for the other to speak Swedish, i.e. to use the One Parent One Language strategy, OPOL/1P1L (see e.g. Barron-Hauwaert, 2004; De Houwer, 2009, pp. 107–113). This was done in 50% of the Swedish-German families, but in only five Swedish-Turkish families (10.6%), which reflects the fact that the Swedish-German group contains many families in which one parent is a native speaker of German and the other parent is a native speaker of Swedish. In both bilingual groups, it was uncommon for both parents to use both languages, and in a small number of families in both groups, both parents spoke primarily Swedish to the child.

In Table 3.9, the child's reported language use with the parents is shown.

[26] For two children in the Swedish-Turkish group, information about parents' language to child and child's language to the parents was only available for one parent. In these two cases, the information from the other parent was used for both. It can be assumed that in these cases, the language use with the other parent was similar, or that there was just one primary caregiver. For one Swedish-Turkish child, information about language use was missing completely.

[27] In one Swedish-Turkish family and two Swedish-German families, one parent spoke the minority language and one parent another language (Kurdish and Chinese, respectively) and in one Swedish-German family, one parent spoke German and the other one both German and English to the child.

Table 3.9. Child's reported language use with parents (%, number of children in parentheses).

	Swedish-Turkish bilinguals (N=48)	**Swedish-German bilinguals (N=46)**
To both almost only or mostly SWE	10.6% (5)	17.4% (8)
To 1 50/50, to 1 almost only/mostly SWE	4.3% (2)	17.4% (8)
OPOL (to 1 almost only/mostly ML, to 1 almost only/mostly SWE)	4.3% (2)	23.9% (11)
To both 50/50	8.5% (4)	2.2% (1)
To 1 mostly/almost only ML, to 1 50/50	6.4% (3)	0% (0)
To both almost only/mostly ML	65.9% (31)	34.7% (16)
Other[28]	0% (0)	4.4% (2)
Total	**100% (47)**	**100% (46)**

Note. SWE = Swedish, ML = minority language (German or Turkish), OPOL = one parent, one language.

The number of children who were reported to speak almost only or mostly the minority language to both parents is more or less identical to the number of children who were reported to receive this type of parental input (cf. Table 3.8). Thus, when both parents consistently use the minority language to the child, the child also uses the minority language. When parents speak the minority language less, the child tends to speak more of the majority language (Swedish). For example, in 16 Swedish-German families, one parent spoke German whereas the other mainly used Swedish with the child, whereas only 11 children were reported to follow this pattern of language use. The other five children spoke less German and more Swedish to both parents. In total, seven Swedish-Turkish children and 16 Swedish-German children were reported to speak mainly Swedish to their parents. This can be compared with only two children in each group receiving such input from their parents.

A final important piece of information about the children's language environment is the approximate time spent in an environment where the languages are spoken. There are some differences between the two groups in terms of exposure patterns, as shown in Table 3.10. Generally, the Swedish-Turkish children received more minority language input in their everyday life than the Swedish-German children, at least in terms of percentage exposure to the two languages. Eight Swedish-Turkish children were reported to receive 80% or more Turkish input, whereas no Swedish-German child heard more than 60% German. Similarly, the number of Swedish-German

[28] These were two Swedish-German families, one in which the child spoke German to one parent and Chinese to the other, and one in which the child spoke German to one parent and German and English to the other. Note that the second Swedish-German child (BiGer5-04), who also received Chinese input from one parent, was reported to mostly use Swedish, and only sometimes Chinese or German with his parents.

children receiving 80% or more Swedish input was higher than in the Swedish-Turkish group (16 Swedish-German children; nine Swedish-Turkish children). However, in both groups, the majority of the children were reported to receive at least 60% Swedish input (32 Swedish-German children; 25 Swedish-Turkish children).[29]

Table 3.10. The child's exposure to the two languages according to parental report (%, number of children in parentheses).

	ML 95% SWE 5%	**ML 80% SWE 20%**	**ML 60% SWE 40%**	**ML 50% SWE 50%**	**ML 40% SWE 60%**	**ML 20% SWE 80%**	**ML 5% SWE 95%**
Swedish-Turkish (N=48)	4.2% (2)	12.5% (6)	27.1% (13)	4.2% (2)	33.3% (16)	18.8% (9)	0% (0)
Swedish-German (N=46)	0% (0)	0% (0)	19.6% (9)	2.2% (1)	34.8% (16)	30.4% (14)	4.4% (2)

Note. ML = the minority language (Turkish or German), SWE = Swedish.

3.2 Elicitation instruments

In this section, the elicitation instruments used in the current study are described. For more details, the reader is referred to manuals and key publications for each stimulus material.

Data was collected from the children using three instruments. All children were tested in both languages with Cross-linguistic Lexical Tasks (CLTs). Two different instruments were used for eliciting three narratives per child in each language, the Multilingual Assessment Instrument for Narratives (MAIN), from which each child told two stories per language, and the Edmonton Narrative Norms Instrument (ENNI), from which each child told one narrative per language.[30] ENNI was chosen in addition to MAIN to be able to compare the children's narrative production on two different elicitation instruments.

3.2.1 Cross-linguistic Lexical Task (CLT)

The Cross-linguistic Lexical Tasks (CLTs) are picture-based vocabulary tasks with four parts: noun comprehension, verb comprehension, noun pro-

[29] Two of the Swedish-Turkish children were reported to hear Kurdish as well as Turkish and Swedish. Additionally, parents of four Swedish-German bilinguals (8.7%) had chosen the option 'other'. In two families, Chinese was also spoken (by the mother). In these cases, 70% and 80% Swedish was heard by the child, and the proportions of Chinese and German were rated to be 15% and 10%. Two further children received input in an additional language, namely English. One child heard 75% Swedish, 20% English, and only 5% German, whereas for the other child there was the same amount of English, but instead 50% Swedish and 30% German.

[30] The Swedish-Turkish bilinguals only told an ENNI-story in Swedish, cf. Section 3.3.

duction and verb production.[31] The CLTs are a set of unique vocabulary tasks in different languages that are based on the same (word)list of basic concepts for objects and actions (corresponding to 158 nouns and 142 verbs) and an accompanying picture database. The goal was to create comparable tasks to assess vocabulary in bilingual children as well as compare results between different languages. What makes the CLT different from other vocabulary tests that have multiple language versions is that the CLTs are not translations but were developed for each individual language from a common base. This means that each language version has been adapted to the linguistic and cultural setting of the country where the language is spoken. The target words are culturally appropriate and have different levels of difficulty in the specific language. Each language version has its own unique combination of target words and distractors. There may also be more than one version for a specific language – for example, since Great Britain and South Africa differ both socio-culturally and environmentally in ways that influence everyday vocabulary, the British and South African varieties of English have their own CLTs. While no two language versions are identical, there is some overlap in the target items, since the same list of words was used to create each version.

The CLTs were designed to be used primarily with children in the late preschool age (3–5/6 years). Scores from monolingual five-year-olds can function as baseline data to which other groups can be compared (Haman, Łuniewska, & Pomiechowska, 2015, p. 220). The level of the vocabulary assessed is such that a monolingual five-year-old should be able to complete the test with a good score. The test scores are therefore not normally distributed, but older (preschool) children should be able to answer most items correctly. This is common for tests developed to measure a child's ability to master tasks at a certain level (cf. Frylmark, 2006). A recent study of monolingual children aged 3–6 (Haman, Łuniewska, et al., 2017), found similar results for 16 out of 17 studied languages (cf. Section 5.1.1), and showed that the test is suitable for this age group, as "it is sensitive to the participants' age and differentiates well between comprehension and production, as well as between nouns and verbs" (Haman, Łuniewska, et al., 2017, p. 19). Previous studies also indicate that typically-developing children and children with (S)LI score differently on the CLT, as has been shown both for monolingual Slovak children (Kapalková & Slančová, 2017) and for Lebanese

[31] They were developed by Working Group 3 Lexical and Phonological Processing of the COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment" funded by European Cooperation in Science and Technology (COST). The action took place 2009-2013 and had the aim of creating a network for working with issues related to disentangling bilingualism from specific language impairment and to create tools for assessing bilingual children's language abilities. For more information about COST Action IS0804, see http://www.bi-sli.org.

bilinguals (Khoury Aouad Saliby, dos Santos, Kouba Hreich, & Messarra, 2017).

The CLTs were specifically constructed with the goal of assessing the vocabulary of bilingual children: "The aim was to design uniform tools for use in bilingual populations of any language pair from all the languages included" (Haman et al., 2015, p. 198). Additionally, the CLT enables comparative studies between different languages. At present, CLT is available for 26 languages, and new language versions are added continuously.[32] The Swedish CLT was constructed and piloted by Gisela Håkansson and Natasha Ringblom together with the author. The German CLT was constructed by Tanja Rinker and Natalia Gagarina, and the Turkish CLT by Özlem Ünal-Logacev, Aylin Müge Tuncer and Pinar Ege.

Each of the four parts consists of 30 test items and 2 practice items. The test thus has 128 items in total, with a maximum total score of 120 points.[33] The two comprehension parts are so-called 'picture choice' or picture-identification tasks, where the child hears the target word embedded in a question (e.g. *Who is digging?* for verb comprehension) and has to point to the picture showing the word. The two production parts are picture-naming tasks, in which the child sees one picture and answers the question from the experimenter (e.g. *What is that?* for noun production) with a word. Each item is presented in landscape orientation on a separate page, in A4-format for the comprehension items and in A5-format for the production items. All pictures are colored. Examples of items are found in Figure 3.1.

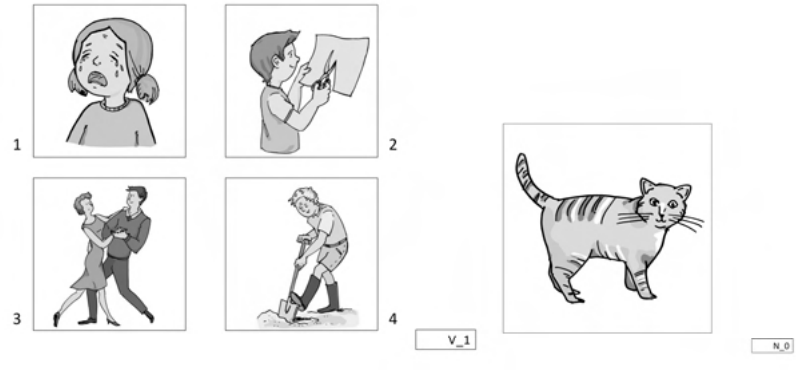

Figure 3.1. Examples of CLT verb comprehension (left) and noun production (right) items (small greyscale copies). Items are taken from the Swedish CLT. Copyright of all CLT pictures: University of Warsaw.

[32] For an updated list of CLT versions available and their authors, see http://psychologia.pl/clts/

[33] Note that in the current study, only results from the production parts were analyzed. For a study of both comprehension and production of the monolingual children and the Swedish-German bilinguals of the current study, see Lindgren (2017) and Lindgren & Bohnacker (submitted), respectively.

In the current study, the CLTs were administered as paper and pencil tasks.[34] The experimenter showed the child the test items in the form of four booklets with laminated pages, asked the questions for all items, and filled in the child's answers on a paper form. If the child did not answer the question for one item, the question was repeated once. If the child still gave no answer, the experimenter wrote down 'no answer' and continued with the next item. Following the standard procedure (Haman et al., 2015, p. 221), the experimenter only provided neutral feedback (*aha*, *okay*) as response to the child's answers and did not answer questions from the child about his/her performance. After finishing the last of the four parts, the child was always praised, irrespective of the actual outcome of the testing.

For a detailed description of the CLTs and how they were constructed, see Haman et al. (2015).

3.2.2 Multilingual Assessment Instrument for Narratives (MAIN)

The Multilingual Assessment Instrument (MAIN) was created to be used with children aged three to ten and consists of two pairs of picture sequences, with one pair containing the stories Cat and Dog (see Figure 3.2) and the second one the stories Baby Birds and Baby Goats (see Figure 3.3). Cat and Dog were created within COST Action IS0804. Baby Birds was based on the 'Cat Story' (Hickmann, 2002) and Baby Goats was loosely based on the 'Fox story' (Gülzow & Gagarina, 2007), but developed further within the COST action.[35] Each picture sequence has six pictures depicting a story containing three episodes. Each episode contains a goal-attempt-outcome sequence for a character (cf. Section 1.1). The picture sequences have been carefully constructed to be comparable following a similar procedure (cf. Gagarina et al., 2012, Section 2.4); all four stories are parallel in terms of length and story grammar components, and the two stories within each pair also contains the same number of characters. MAIN "has advantages over longer and more elaborate narrative elicitation methods in that it is carefully structured, allowing identification of the category that has been generated or retold by the child" (Gagarina et al., 2015, p. 255). It also provides the child with more opportunities to produce a specific macrostructural component and/or sequence (cf. Sections 1.1 and 7.3), compared with shorter, single-episode stories. MAIN is suitable for studying bilinguals' narratives as it "allows for the evaluation of narrative skills in dual languages of bilingual children using similar stimuli for these languages" (Gagarina, Klop, Tsimpli, & Walters, 2016, p. 15).

[34] Some language versions, including German, are also available as computerized versions.
[35] MAIN was developed by Working Group 2 Narrative and Discourse of the aforementioned COST Action IS0804.

The MAIN picture sequences, presented to the child in color as a fold-out strip, can be used to elicit narratives in three different modes: *telling* (story generation), in which the child, after having looked at the pictures, simply tells the story shown in the pictures, *retelling*, in which the child retells the story s/he heard from an adult with support of the pictures, or *model story*, in which the child first hears one of the stories from an adult while looking at the pictures, and then tells another one of the stories. In all three different modes of elicitation, comprehension questions are also asked after the story. For each story, there are ten comprehension questions, specifically targeting comprehension of characters' goals and internal states. Although Cat and Dog were originally created to be used in the retelling mode,[36] in the current study, all four picture sequences were used for telling. For the rationale behind this choice, see Section 1.4.1. The reason for including all four picture sequences instead of only Baby Birds and Baby Goats was to collect more data from each child and to be able to compare their performance on the two tasks (Cat/Dog vs Baby Birds/Baby Goats).

In addition to the picture sequences and the comprehension questions, the MAIN includes a protocol for assessing production of macrostructural elements (or story structure components) and narrative (structural) complexity, which has been used in the current study.[37] Following this protocol, the children are awarded points for including the setting (time and place) of the story, and characters' goals, attempts and outcomes, as well as internal states as initiating events (i.e. forming the starting point of an episode) and as reactions (i.e. the characters' emotional reaction to the episode's outcome) for each of the three episodes. The method of analyzing comprehension and production of macrostructure according to the MAIN protocol is described in detail in Sections 7.2.1 and Section 7.3.1, respectively. The standard prompts and comprehension questions from the Swedish, German and Turkish versions of the MAIN were used.

The Cat and Dog stories (Figure 3.2) have identical plotlines but contain different characters and objects. The number of characters is the same (three characters). There is a cat/dog that wants to catch a butterfly/mouse, jumps to try to catch it, but lands in the bush/tree (episode 1). Simultaneously, a boy comes. He loses his ball/balloon in the lake/tree, wants to get it back, attempts to do so and then retrieves it (episode 2). Meanwhile, the cat/dog

[36] Although created for retelling, the construction process of Cat/Dog did not differ from that of Baby Birds/Baby Goats (cf. Gagarina et al., 2012, Section 2.4); in all cases the picture sequences were created first and scripts for retelling (which do exist for Baby Birds/Baby Goats as well) were created at a later stage. The assessment of narrative macrostructure in comprehension and production, following the MAIN manual (see Chapter 7), is identical for all stories irrespective of elicitation mode.

[37] Additionally, MAIN includes a section for scoring internal state terms (ISTs), i.e. word tokens used to express the emotions (e.g. *happy*, *sad*), perceptions (e.g. *seeing*) and cognitive activities (e.g. *thinking*) of the story characters. This aspect has not been analyzed in the current study.

sees the fish/sausages that the boy brought, takes them and eats them (episode 3). For the full story scripts of Cat and Dog, see Gagarina et al. (2012, p. 139).

The Baby Birds and Baby Goats stories (Figure 3.3) are also parallel in terms of episodic structure and contain the same number of characters (five characters). However, they are not as identical as Cat and Dog are. The difference between the picture sequences concerns the content of episode 1. In Baby Birds, the baby birds are hungry so the mother/parent bird flies away to get food and then feeds them, whereas in Baby Goats, one of the baby goats is drowning and the mother rescues it (whilst the other baby goat is eating grass). Episodes 2 and 3 are more or less the same, with a 'baddie', the cat/fox, trying to catch the baby bird/goat (episode 2) and a 'hero', the dog/bird, coming to the rescue (episode 3). For the story scripts of Baby Birds and Baby Goats, see Gagarina et al. (2012, p. 138).

The standard procedure of the MAIN was followed. The experimenter is not allowed to look at the pictures before or during the storytelling and acts as if the stories are unknown to him/her. Three colored envelopes are placed on a table. The child is asked to choose one envelope and take out the pictures inside. All envelopes contain the same story, but in this way, the child is led to believe that the experimenter does not know which story the child is going to tell. The child takes out the pictures and looks at all the pictures before telling the story. When the child has finished looking at the pictures, the experimenter folds them back so that only pictures 1-2 are visible to the child (but not to the experimenter), and asks the child to begin telling the story. When the child has finished telling about the first two pictures, the next two pictures (pictures 3-4) are unfolded and finally the last two (pictures 5-6), so that all six pictures are visible to the child. When the child shows signs of having finished the story, the experimenter asks if s/he is finished. On confirmation, the pictures are placed on the table so that they are visible to both child and experimenter. The experimenter then proceeds to ask the comprehension questions. Throughout the storytelling, the experimenter is supposed to give only minimal prompts (e.g. *aa*, *mm*, *and then*?). In the current study, the younger children generally required more frequent prompting than the older ones, but, regardless of their age, all children who needed prompting were supposed to receive the same types of prompts. All children received praise (e.g. *Vilken fin saga*! 'What a nice story!') after telling the story.

For more information about the MAIN, its procedure and how it was developed, see Gagarina et al. (2012, 2015).[38]

[38] MAIN can be downloaded from http://www.zas.gwz-berlin.de/zaspil56.html after registering as a user.

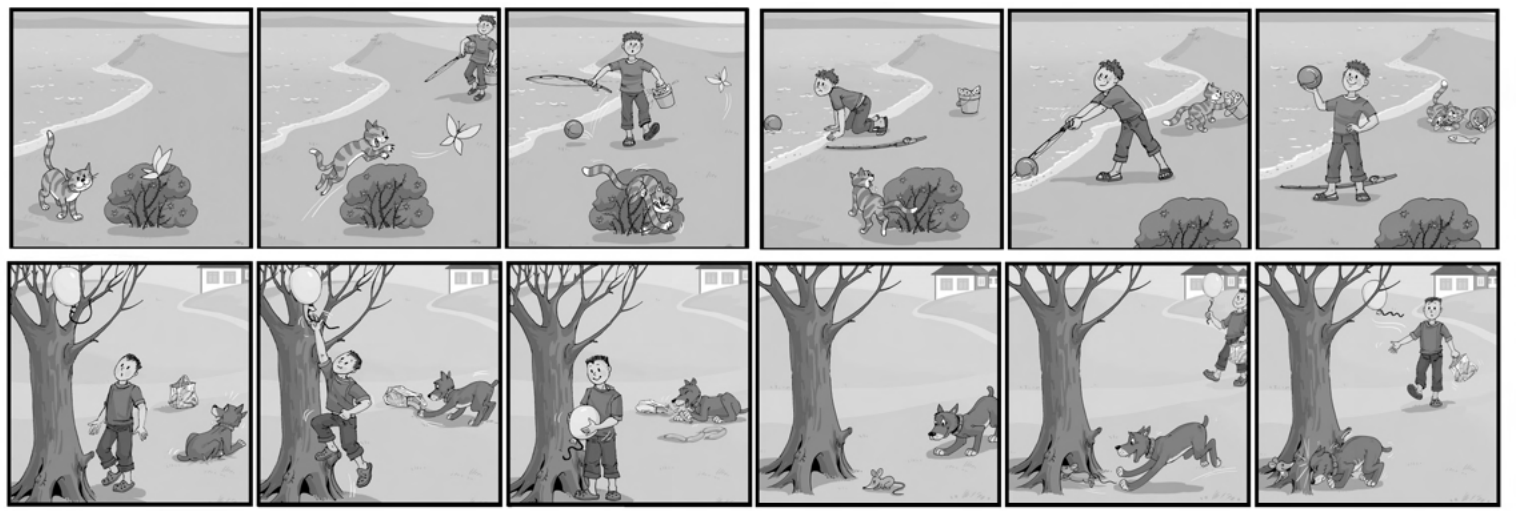

Figure 3.2. Small-scale copies of the picture sequences Cat (above) and Dog (below) from the Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012, 2015).

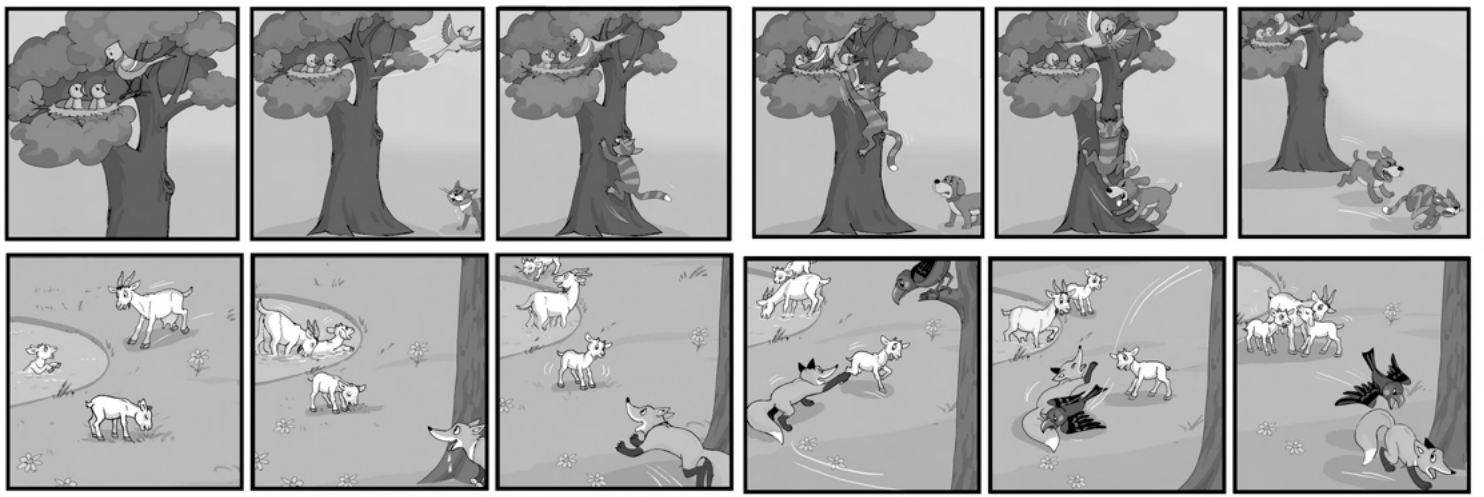

Figure 3.3. Small-scale copies of the picture sequences Baby Birds (above) and Baby Goats (below) from the Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012, 2015).

3.2.3 Edmonton Narrative Norms Instrument (ENNI)

The Edmonton Narrative Norms Instrument (ENNI) is a narrative assessment instrument developed for children between age 4 and 9 (Schneider et al., 2005). The complete ENNI consists of two sets of three stories, which were constructed to be parallel in order to allow testing of bilinguals' both languages. The three stories in each set have different levels of complexity. According to the original ENNI assessment protocol, each child tells all three stories. In the current study, only the two stories (one from each set) with the medium level of complexity were used, A2 and B2 (see below for rationale). Both A2 and B2 consist of eight pictures depicting a story with three characters, which are all humanized animals. Both contain two episodes with a complication/problem for one character and help provided by other characters. The pictures are black-and-white line drawings, in a 'cartoonish style' (see Figure 3.4).[39]

In A2, the setting is a swimming pool and the characters are an elephant, a giraffe and an (elephant) lifeguard. The elephant is standing next to the pool with the giraffe. There is a sign showing that no running is allowed. The elephant wants to run to jump into the pool. She runs and then falls and hurts her knee. The giraffe runs after her to help her, and the lifeguard comes to see what has happened. He puts a band aid on the elephant's knee and they help her to sit down on a bench. In the final picture, he points to the no running sign with a serious expression; the elephant looks ashamed.

The B2-story takes place in the park and the characters are a rabbit, a dog and a (rabbit) doctor. The rabbit comes walking in the forest towards the dog. Both are carrying picnic baskets. They sit down to have a picnic and the rabbit starts to eat a lot of food that he has brought. He eats all the food and gets a stomach ache and feels dizzy. The dog runs to get help, finds a doctor and convinces her to come and have a look at the rabbit. She examines him and then brings him with her, while the dog remains standing next to the picnic blanket.

Figure 3.4 shows small-scale copies of picture 1, in which the first two characters are present, and picture 5, in which the third character enters the story, for A2 and B2, respectively.

[39] ENNI has been norm-referenced for English for a number of different types of analyses, e.g. Story Grammar, Syntactic Complexity Measures and First Mentions (using a scoring system which includes characters and objects in all three stories of each set).

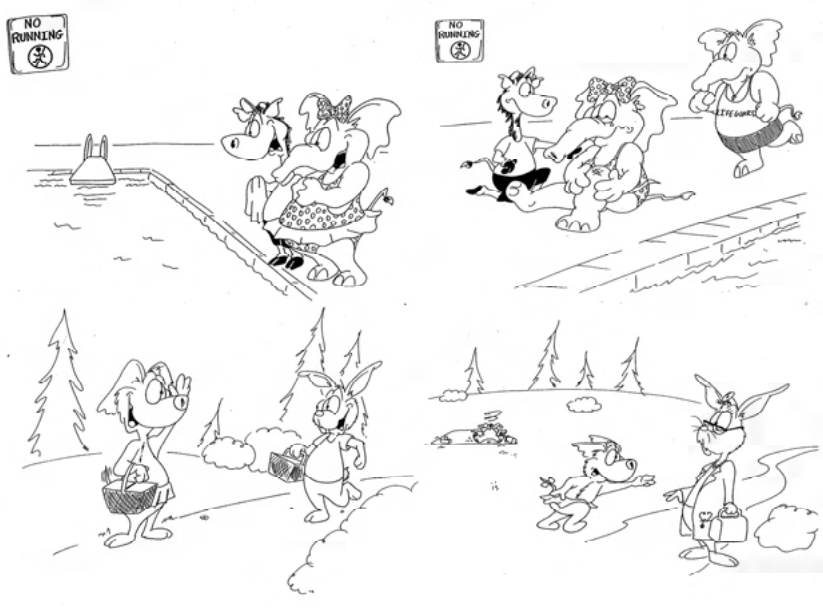

Figure 3.4. Small-scale copies of pictures 1 and 5 from A2 (above) and B2 (below) of the Edmonton Narrative Norms Instrument (Schneider et al., 2005). ©2000, Wooket Graphics. (Reprinted with permission).

In the current study, the pictures were laminated, put together and presented to the child as a booklet in A5 format, with one picture on each page. The ENNI procedure, in which the experimenter holds the pictures so that only the child can see them, was followed closely. Before telling the story, the child was allowed to look at all pictures. Narratives were only elicited with ENNI in Swedish and German (cf. Section 3.3). The instructions and prompts were translated from English (see Schneider et al., 2005) to Swedish by the author. The German translation of the instructions and prompts was done by the author together with a native-speaker research assistant and checked by a native speaker linguist.

ENNI was chosen in addition to MAIN to be able to compare data from the same children on two different elicitation instruments, in order to answer research question 4 about task effects. In the current study, MAIN and ENNI were only compared for character introduction (see Chapter 6). The ENNI A2 and B2 stories were chosen because they contain the same number of characters as the MAIN Cat/Dog, thus making it possible to compare character introduction in these two narrative tasks. MAIN and ENNI differ in several respects, both concerning aspects of story grammar (i.e. number of episodes) and types of characters, and visual aspects, such as the drawing style. Table 3.11 gives an overview of similarities and differences between MAIN1 Cat/Dog and ENNI A2/B2 both in terms of general aspects, such as

number of pictures/episodes and type of story presentations, and aspects related to the story characters, such as when they enter the story.[40]

Table 3.11. Similarities and differences between MAIN Cat/Dog and ENNI A2/B2. Table from Lindgren (2018).

	MAIN1 (Cat/Dog)	**ENNI (A2/B2)**
Type of story presentation	Fold-out picture sequence	Booklet
Color and style	Full color	Black-and-white line drawing
Number of pictures	6 pictures	8 pictures
Number of episodes	3 episodes	2 episodes
Number/types of characters	3 characters; 1 animate agent, 1 animate non-agent, 1 human	3 characters, all humanized animal agents
Main vs auxiliary characters	2 main characters	1 main character
Characters enter when?	Characters 1 and 2 in picture 1, Character 3 in the background of picture 2	Characters 1 and 2 in picture 1, Character 3 in picture 5

3.3 Setup and procedure

In this section, the general data collection procedure is described, including the counterbalancing systems used (Section 3.3.1) and a summary of the pilot study carried out to test the procedure (Section 3.3.2). Standard procedures for administering the different tasks to the children were followed (see Section 3.2).

With all children the same general procedure was followed. Data from the two languages of the bilingual children were collected on separate occasions. Half of the children were tested in Swedish first and half in German/Turkish first. The aim was to have 5–7 days between the testings. Due to unforeseen events, such as a child being sick, or general difficulties with scheduling the testings, e.g. due to school holidays, this was not always possible. For the Swedish-German participants, the time between testings varied between 2 and 21 days, with an average time of 8.7 days (SD = 3.7). For the Swedish-Turkish bilinguals, the time between testings varied between 4 and 31 days, with an average time of 12.8 days (SD = 6.4). For 65% of the Swedish-German bilinguals and 44% of the Swedish-Turkish bilinguals the time between the testings was 3–10 days.

[40] For a study comparing character introductions in MAIN (Cat/Dog) and ENNI (A2/B2) by the same Swedish monolinguals as in the current study, see Lindgren (2018). In Lindgren (2018), the potential effects of the aspects presented in Table 3.11 on character introductions are discussed.

All monolingual data and the Swedish data from the Swedish-German bilinguals were collected by the author, a native speaker of Swedish. The German data were collected by Valerie Reichardt and Ute Bohnacker, both native speakers of German. The data from the Swedish-Turkish bilinguals were collected within the BiLI-TAS project (Bohnacker, 2013). In total, six different experimenters, including research assistants and senior researchers, as well as the author herself, collected Swedish data from the Swedish-Turkish children participating in the present study.[41] All experimenters who tested the children in Swedish were native or near-native speakers of Swedish. Three different experimenters, Buket Öztekin and two research assistants, all native speakers of Turkish, tested the children in Turkish.

Data collection took place with each child individually in a quiet room at the (pre)school or at home (bilinguals only). Three Swedish-German bilinguals were tested at Uppsala University. The experimenter spoke to the child only in the language of the testing, and acted as if s/he did not understand the other language. This was done to create a monolingual testing situation in order to encourage the child to use the language of the testing as much as possible. In most cases, the experimenter, together with a preschool teacher (or when the testing took place at home, a parent), met the child outside the room where the data collection was conducted. After introducing herself, the experimenter went with the child to the room. In some cases, the materials and recording equipment were put in order before the child entered the room; in other cases, the experimenter set it up while chatting with the child. The child and the experimenter sat at opposite sides of a table, so that the experimenter could not see the narrative stimuli. The envelopes with the picture sequences for the first story from MAIN were, whenever possible, placed on the table before the child entered the room. After a short warming-up phase in which the experimenter asked the child some general questions about e.g. spare time activities, siblings, pets etc.,[42] the experimenter explained that the child was going to look at some pictures, tell some stories, and answer some questions. The experimenter then proceeded with the testing, according to the pre-determined order of the tasks (see Section 3.3.1) and following the scripts of the respective elicitation instruments closely. All children thus received the same instructions for each of the tasks. Data collection took approximately 25–45 minutes.

All sessions were both video and audio recorded. The video recorder was running throughout the entire session, including the warming-up phase. Additionally, a digital voice recorder was used to audio record either the whole session (in which case it was turned on after the warming-up phase) or only the narratives, including the answers on the comprehension questions for

[41] The author collected Swedish data from six of the 48 Swedish-Turkish children who participated in the current study.
[42] No script was used for the warming-up phase.

MAIN, and the production parts of the CLT. The audio recorder was always placed on the table near the child.

The video recorder was, whenever practically possible, set up so that the narrative stimuli as well as at least a part of the child's face were visible, e.g. with the video recorder positioned in the corner somewhat behind the child. This ideal situation is shown in Figure 3.5. Experimenters tried to get as close as possible to the ideal setup, but in many cases the size and shape of the room and/or table meant that the camera could not be placed as in Figure 3.5. Therefore, in some recordings, the narrative stimuli are not fully visible.

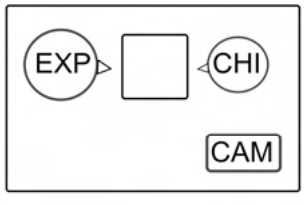

Figure 3.5. The ideal placement of the experimenter (EXP), the child (CHI) and the video recorder (CAM).

In Swedish and German, all children first told either Cat or Dog (MAIN1), followed by the CLT in the order that had been assigned to the child (see Section 3.3.1), then either Baby Birds or Baby Goats (MAIN2) and finally the A2 or the B2 (ENNI). This is schematized in Figure 3.6 below. After each MAIN narrative, the child was asked the comprehension questions. The structure was identical in both languages of the Swedish-German bilinguals.

Figure 3.6. Structural overview of the data collection, Swedish and German. MAIN = Multilingual Assessment Instrument for Narratives, CLT = Cross-linguistic Lexical Task, ENNI = Edmonton Narrative Norms Instrument.

As shown in Figure 3.7, the procedure in Turkish was slightly different from that of Swedish and German. The reason was that the Swedish-Turkish bilinguals were a part of the BiLI-TAS project (Bohnacker, 2013) which also included non-word repetition (NWR) tasks for screening purposes. Instead of the ENNI, in Turkish, the Swedish-Turkish bilinguals therefore did two NWR tasks, one Quasi-universal and one Turkish NWR (Chiat, 2015; Topbaş & Kaçar, 2013; Topbaş, Kaçar-Kütükçü, & Kopkalli-Yavuz, 2014), one at the beginning and one at the end of the session.[43]

[43] The data from these NWR tasks are analyzed in Öztekin (in preparation).

Figure 3.7. Structural overview of the data collection, Turkish. NWR = non-word repetition task, MAIN = Multilingual Assessment Instrument for Narratives, CLT = Cross-linguistic Lexical Task.

Between each task there was a short break in which the child was given a sticker (i.e. 4–5 stickers per session). The children received a certificate after completing their participation in the study.

3.3.1 Counterbalancing

In order to ensure that the combinations of stories would occur equally often across all children, counterbalancing of the different tasks was done on the basis of both age and language group, i.e. within each age group for the three language groups separately. Any deviances from the counterbalancing system of the narrative tasks are described and explained in Chapter 4. The counterbalancing system of the narrative tasks for the monolinguals is shown in Table 3.12.

Table 3.12. Counterbalancing of the narrative tasks, monolinguals.

Child	MAIN1	MAIN2	ENNI
1	Cat	Baby Birds	A2
2	Cat	Baby Goats	B2
3	Dog	Baby Birds	A2
4	Dog	Baby Goat	B2
5	Cat	Baby Birds	B2
6	Cat	Baby Goats	A2
7	Dog	Baby Birds	B2
8	Dog	Baby Goats	A2

Note. The system was repeated after every 8th child.

As mentioned above, half of the bilingual children were tested in Swedish first and the other half in German/Turkish first. The bilingual children always received different stories in their two languages, since there were two different stories for each narrative task (MAIN1, MAIN2, and ENNI). The counterbalancing system for the three narrative tasks for the Swedish German bilinguals is shown in Table 3.13. To give an example, a Swedish-German child with the combination Cat-Baby Birds-A2 in Swedish told Dog-Baby Goats-B2 in German. The counterbalancing system for the Swedish-Turkish bilinguals was identical, except that they did not tell an ENNI story in Turkish.

Table 3.13. Counterbalancing of the narrative tasks, Swedish-German bilinguals.

	Language A			Language B		
Child	**MAIN1**	**MAIN2**	**ENNI**	**MAIN1**	**MAIN2**	**ENNI**
1	Cat	Baby Birds	A2	Dog	Baby Goats	B2
2	Cat	Baby Goats	B2	Dog	Baby Birds	A2
3	Dog	Baby Birds	A2	Cat	Baby Goats	B2
4	Dog	Baby Goats	B2	Cat	Baby Birds	A2
5	Cat	Baby Birds	B2	Dog	Baby Goats	A2
6	Cat	Baby Goats	A2	Dog	Baby Birds	B2
7	Dog	Baby Birds	B2	Cat	Baby Goats	A2
8	Dog	Baby Goats	A2	Cat	Baby Birds	B2

Note. The system was repeated after every 8th child.

For the CLT, counterbalancing was as follows. As described above (Section 3.2.1), each CLT consists of four parts, noun comprehension, verb comprehension, noun production, and verb production. The four possible orders of the CLT parts were counterbalanced for each group, as shown in Table 3.14 (for details, see Haman et al., 2015). This means that one fourth of the children received the parts in order 1, one fourth received order 2, and so forth. The bilingual children received the different parts of the CLT in the same order in both languages, so as to make the conditions of the lexical task identical in the child's two languages.

Table 3.14. Counterbalancing Cross-linguistic Lexical Tasks (CLTs).

Order	**Part 1**	**Part 2**	**Part 3**	**Part 4**
1	Noun Comp	Verb Comp	Noun Prod	Verb Prod
2	Verb Comp	Noun Comp	Verb Prod	Noun Prod
3	Noun Prod	Verb Prod	Noun Comp	Verb Comp
4	Verb Prod	Noun Prod	Verb Comp	Noun Comp

3.3.2 Pilot study

In October 2013, a pilot study was conducted with nine Swedish monolinguals (age range 3;10–5;9, 6 girls). All children were recruited from the same preschool in a larger Swedish city. The main aims were to make sure that the procedure and the tasks were suitable for the intended age group, i.e. that the length and the combination of tasks was not too demanding for the youngest children, and to check if the resulting data were rich enough for the analyses. A further aim of the pilot study was for the author to gain experience in working with children of this age as well as in contacting preschool directors, personnel and parents. The pilot study was also part of piloting the Swedish CLT. No part of the pilot data was included in the present study.

In the pilot study, each child told two narratives, first one from MAIN (Cat or Baby Goats) and, after the CLT, either a second narrative from MAIN or ENNI (in all cases the B2 was used). Overall, the procedure

worked well. Data collection took 25–40 minutes and even the youngest children did not experience difficulties in completing the tasks. After evaluating the pilot results, it was judged valuable to include two MAIN narratives as well as one from ENNI in the study proper. As this would not considerably lengthen the time spent with each child considerably (generally, completing ENNI only took a few minutes), it was decided that the final setup would include three narrative tasks, in the order described above. Cat was found to be somewhat easier for the children to tell than Baby Goats and B2, and it was therefore decided to begin the session with Cat/Dog.

3.4 Transcription

Before describing the transcription process for the narrative data, a few words should be said about the CLT production responses. All responses were written down by the experimenter on the test sheets. In most cases, the experimenter later checked with the audio-recordings that everything had been correctly written down. All unclear responses were rechecked by native or near-native speakers of the language who also spoke the child's other language.

All narratives, including answers to the MAIN comprehension questions, were transcribed verbatim.[44] This also included words and feedback signals uttered by the experimenter. Most narratives were transcribed from the audio recordings, using the video recording only in cases where the child's speech was especially difficult to hear and for adding relevant non-verbal information. In a few cases, due to technical difficulties, no audio recording was available or only parts of the narrative had been audio recorded. In these cases, transcription was done directly from the video recording.

Transcription was done by experienced transcribers using a careful procedure which included making several passes over each segment of the recording, letting the transcript rest and then listening to the recording again. In most cases, transcription was first done using the transcription program EXMARaLDA and then exported to the CHAT-format[45] to be used with the program CLAN[46] (MacWhinney, 2000). Before exporting the file, the transcription was always carefully checked against the video recording and non-verbal information such as (relevant) gestures and movements was added to the transcript. The CHAT-format transcriptions were then checked again against the video and audio recordings. In a few cases, due to technical error, no video recording was available.

[44] Transcription was done to enable automatic lexical searches and to do (manual) narrative and syntactic analysis. Phonological analyses were not planned and therefore no phonological/phonetic transcription was carried out.
[45] CHAT stands for Codes for the Human Analysis of Transcripts.
[46] CLAN stands for Computerized Language ANalysis.

Detailed transcription guidelines (Guidelines for the transcription of MAIN, version March 2018) were developed within the BiLI-TAS project (Bohnacker, 2013). These guidelines, based on standard conventions of the CHAT-format (MacWhinney, 2000), were followed in the current study. Some transcription conventions deserve specific mention here.

First, as it is often difficult to determine the exact boundaries of an utterance (cf. Turell & Moyer, 2008, p. 201), it was decided to use the conversational turn as the unit of transcription. This meant that "everything a speaker says before the other speaker talks is generally transcribed as one unit" (Guidelines for the transcription of MAIN, version March 2018, p. 6). Although often done in studies of children's narratives, it was not deemed necessary for the purpose of the current study to separate the turns into so-called c-units.[47]

Second, the length of the pauses were not measured, but pauses were transcribed with (.) for a short pause, (..) for a somewhat longer pause, and (…) for a very long pause, following the CHAT-conventions.

Third, the child's speech was transcribed using standard orthography of the three languages (Swedish, German, and Turkish) unless the pronunciation deviated considerably from standard pronunciation.[48] In these cases, an approximation of the child's pronunciation was used. Words that could not be identified were transcribed as xx, and were counted towards the total word tokens.

Fourth, non-verbal information deemed potentially relevant for our analyses was added to the transcripts. This included pointing, unfolding of the pictures, and actions replacing speech, such as nodding or gestures.

Fifth, commentary was added if the child's speech was generally unclear or contained many phonological simplifications or in other ways differed from the target, or if the sound quality of the recording was very low, e.g. due to background noise.

Sixth, repetitions and re-phrasings were coded with [/], [//] and [///] according to the CHAT-format. Fillers, interrupted/unfinished words and sounds (e.g. imitating noises make by story characters) were coded using the &-symbol. Additionally, all the yes/no-answers given by the child to questions from the experimenter that were not related to the narrative were also coded with & (e.g. the child's affirmative answer to the experimenter's question if the child had finished the narrative). All calculations of number of

[47] A c-unit is similar to a t-unit, but is adapted to the elliptical nature of speech (Loban, 1963). A t-unit is a unit that contains "one main clause with all subordinate clauses attached to it" (Hunt, 1965, p. 20). It was first developed for analyzing writing. For a criticism of the vagueness of the concepts of both t-unit and c-unit, see Foster, Tonkyn & Wigglesworth (2000).

[48] The guiding principle was: "As long as the word is recognisable, do not attempt to transcribe approximations of the spoken form, but use the conventional spelling of the informal written language" (Guidelines for the transcription of MAIN, version March 2018, p. 17).

word types and tokens automatically excluded strings of words marked by [/], [//], [///] and &.

Finally, words in another language (e.g. Swedish words in the German narratives), were marked as code-switches, using the codes [@s], [@g] and [@t] for Swedish, German and Turkish material. This enabled an automatic CLAN analysis of word types and tokens in the narratives including or excluding lexical material from the other language (see Chapter 4).

All monolingual narratives were transcribed by the author. The transcribed narratives of nine monolingual children (12.5% of the monolingual transcriptions) were checked against the audio files by a Swedish native speaker SLP and PhD student of Linguistics (Linnéa Öberg).[49] Any differences heard were noted. The agreement rate between the transcriber and the checker was 98.5%. The differences found were discussed until consensus was reached. When necessary, the audio and/or video files were consulted.

The Swedish narratives of the Swedish-German bilinguals were transcribed by the author and the German narratives by a research assistant who was a native German speaker and a trained linguist, but who spoke no Swedish (Valerie Reichardt). All German transcripts were afterwards checked by the author, a fluent speaker of German, to ensure that transcription conventions were followed and that Swedish lexical material, if present in the child's production, was correctly transcribed. Whenever the transcripts contained (suspected) Swedish words or more than one word that the research assistant had not been able to hear (marked by xx in the transcripts), the author listened to the audio files and made changes whenever she heard something different.[50]

Swedish data from six Swedish-German bilinguals (13% of the transcriptions) were checked by a German native speaker SLP (Sibylle Dillström), and German data from eleven Swedish-German children (24% of the German transcriptions) were checked by a German native speaker linguist (Ute Bohnacker). Both checkers were also near-native speakers of Swedish. The choice of the Swedish-German children to be checked was not random, but based on the child having unclear speech or non-target pronunciation, and/or (mostly in the case of German) weak language proficiency with a high amount of code-switching and/or mixed word forms. There was some overlap between the children checked in Swedish and German, as some children had unclear speech in both languages. The agreement rate between tran-

[49] A random number generator (https://www.random.org/) was used to select the codes of the children whose transcripts were to be checked.

[50] The German narratives of one child (BiGer6-11) were treated differently from the rest. This child spoke Swiss German and the sound quality of the recordings was low, with a lot of background noise, making it especially difficult to transcribe. These narratives were first transcribed roughly by the German research assistant and improved upon by the author, transcribed further by another research assistant who was a fluent speaker of German and also an SLP (Karin Koltay), and finally checked by a linguist who was a native speaker of a southern German variety (Ute Bohnacker).

scripts and the checker's judgment was 96.6% for Swedish and 96.5% for German. All differences between transcriber and checker were discussed until agreement was reached.

The Swedish MAIN narratives of the Swedish-Turkish bilinguals were transcribed by one researcher (Sibylle Dillström) and two research assistants within the BiLI-TAS project (Bohnacker, 2013). The ENNI narratives of this group were transcribed by the author. Swedish MAIN and ENNI narratives from three randomly selected children in each age group from the BiLI-TAS projects' larger data set of 102 children aged 4–7 were checked by a Swedish native trained SLP research assistant (Karin Koltay). For the four- to six-year-olds, this resulted in an agreement rate of 98.8%. All differences were resolved through discussion between transcriber and checker.

Raw transcripts of the Turkish MAIN narratives were made by a native Turkish research assistant. The transcripts were later finalized by a native Turkish SLP and PhD candidate (Buket Öztekin). The narratives of 16 four- to six-year-old children (around five children in each age group) of the Bi-LITAS project were then checked by the same research assistant who did the raw transcripts. Some children were selected to be checked because their speech had been especially difficult to transcribe and some were randomly selected. All cases when there was disagreement were checked again against the audio files by the transcriber. The agreement rate was 99.3%.

A final check to ensure that the transcripts followed the transcription conventions was carried out by the author for Swedish and German and by Buket Öztekin for Turkish.

To summarize, transcription was carried out using a carefully designed and consistent procedure that showed high agreement rates between the original transcripts and an independent checker. Examples of transcribed narratives elicited with MAIN Cat/Dog and Baby Birds/Baby Goats can be found in Appendix 1.

4 Narrative production data

As explained above (Section 3.3), all children told three stories in Swedish: MAIN1 (Cat or Dog), MAIN2 (Baby Birds or Baby Goats), and ENNI (A2 or B2). The Swedish-German bilinguals also told three stories in German, whereas, in Turkish, the Swedish-Turkish bilinguals only told the two MAIN-stories. This chapter describes the corpus of narrative production data that has been collected and analyzed.[51] The number of narratives for each narrative task (Section 4.1), the total number of word tokens (Section 4.2), and the mean number of words per narrative (Section 4.3) are reported for Swedish, German, and Turkish.

4.1 Number of narratives

Table 4.1 shows the number of narratives in the Swedish, German and Turkish.

Table 4.1. Number of narratives in Swedish, German and Turkish.

	MAIN1 (Cat/Dog)	**MAIN2 (BB/BG)**	**ENNI (A2/B2)**	**Total**
Swedish	**165**	**166**	**161**	**492**
Monolinguals (N=72)	72 (36/36)	72 (36/36)	72 (36/36)	216
Swedish-German bilinguals (N=46)	45 (22/23)	46 (22/24)	45 (24/21)	136
Swedish-Turkish bilinguals (N=48)	48 (24/24)	48 (24/24)	44 (21/23)	140
German				
Swedish-German bilinguals (N=46)	**46** (23/23)	**46** (24/22)	**44** (21/23)	**136**
Turkish				
Swedish-Turkish bilinguals (N=48)	**48** (24/24)	**48** (24/24)	-	**96**

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

[51] The data from narrative comprehension, i.e. the answers the MAIN comprehension questions, is described in Section 7.2.1.

The data consist of 492 narratives in Swedish, 136 narratives in German, and 96 narratives in Turkish.[52] All monolingual data were collected in accordance with the counterbalancing system used in the study (see Section 3.3.1). In each monolingual age group, there were 72 narratives, evenly divided by narrative task (MAIN1, MAIN2, and ENNI). Each specific story (e.g. Cat) was told by a total of 36 children, i.e. 12 children in each age group.

For the bilinguals, there were some smaller divergences from the counterbalancing system (see Section 3.3.1). One bilingual Swedish-German five-year-old (BiGer5-14) started to tell the first MAIN-story (Cat) in Swedish but then switched to German. The experimenter did not manage to get the child to switch back to Swedish. The child also answered the comprehension questions of this story in German. This narrative, including the answers to the comprehension questions, is not included in the data.

Due to the number of Swedish-German four-year-olds that had to be excluded from the study (see Section 3.1.2), there are two more Baby Goats-narratives than Baby Birds' in Swedish and vice versa in German. In the Swedish testing, one Swedish-German four-year-old (BiGer4-09) did not say anything except 'I do not want' and 'I do not know' in response to the ENNI-pictures, and her production was therefore excluded from the data. Because of this and because a number of children were excluded from the study, there are eight A2-narratives, but only five B2-narratives in the Swedish-German four-year-olds in Swedish, and vice versa in German. Two Swedish-German bilinguals (BiGer4-14, BiGer6-02), received the wrong ENNI-story as part of their German testing, i.e. they were asked to tell the same story they had already told in Swedish. The resulting German narratives were excluded from the data, as they would not be comparable to the German narratives from the other children.

Three Swedish-Turkish four-year-olds (BiTur4-06, BiTur4-07, BiTur4-14) and one five-year-old (BiTur5-21) did not tell any ENNI-story. Due to experimenter error, one five-year-old and one six-year-old did not receive the correct ENNI-story according to the counterbalancing system. This means that there is an uneven number of A2 and B2 narratives for these age groups (see Appendix 2, Table A2.1).

4.2 Number of words

In addition to the number of narratives analyzed, an overview of the size of the corpus in word tokens is needed. The number of word tokens was calculated for each narrative and child using the frequency (freq) analysis in CLAN. For the bilinguals' narratives, two analyses were carried out: all tokens irrespective of language and only tokens in the language of the testing.

[52] For an overview by age group within each language group, see Appendix 2, Table A2.1.

Here, the total number of word tokens per language group is reported; for figures by age group, see Appendix 2, Tables A2.2-A2.4.

Table 4.2 shows the total number of word tokens produced by each of the language groups in their Swedish narratives.

Table 4.2. Total number of words in the Swedish narratives, by language group and narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)	Total
Monolinguals (N=72)	5,676	4,877	5,822	16,375
Swedish-German bilinguals (N=46)	3,432	3,167	3,380	9,979
Swedish words	3,420	3,152	3,372	9,944
German words	12	15	8	35
Swedish-Turkish bilinguals (N=48)	3,637	3,859	4,144	11,640
Total	**12,735**	**11,888**	**13,338**	**37,961**

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

As can be seen in Table 4.2, a few German words were found in the Swedish narratives of the Swedish-German bilinguals, corresponding to 0.4% of the data. Code-switches to German were thus uncommon in the Swedish data. One child (BiGer4-16) produced 60% of the German words (21 words). No other child produced more than four German words in their Swedish narrative. No Turkish words were found in the Swedish-Turkish data; it is possible that some of the words the transcribers were unable to transcribe (i.e. words transcribed as xx) were in fact Turkish.

Table 4.3 shows the total number of words in the German narratives, by narrative task. The number of Swedish words found in the German narratives was considerably higher than the number of German words in the Swedish data, but still made up only 2.3% of the data. Half of the children (23 children) used at least one Swedish word, and 15 of them did so in more than one narrative. The number of Swedish words per child ranged from one word (four children) to 57 words (one child). Ten children were responsible for 89% of the Swedish words. In fact, two children together (BiGer4-15, BiGer5-13) produced 38% of all Swedish words (108 words). This shows that the use of a high number of Swedish words in the German narratives is mainly a phenomenon restricted to a few of the children in the study.

Table 4.3. Total number of words in the German narratives, Swedish-German bilinguals (N=46), by narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)	Total
German words	4,500	3,438	4,262	12,200
Swedish words	82	76	124	282
Total	**4,582**	**3,514**	**4,386**	**12,482**

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

In Table 4.4, the total numbers of words in the Turkish narratives are shown. Swedish words constituted 1.6% of the Turkish narratives, a considerably smaller proportion than in the German narratives. Only 16 children produced one or more Swedish words. Three children (BiTur4-26, BiTur5-10, BiTur5-21) produced 64% of the Swedish words. No other child produced more than five Swedish words. It is thus clear that, in the Swedish-Turkish group, only a few children used many code-switches in their Turkish narratives.

Table 4.4. Total number of words in the Turkish narratives, Swedish-Turkish bilinguals (N=48), by narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	Total
Turkish words	2,919	3,236	6,155
Swedish words	46	51	97
Total	**2,965**	**3,287**	**6,252**

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

To summarize, the corpus of oral narratives analyzed in the present study consists of 37,961 word tokens in the Swedish narratives (out of which 33 words were German), 12,482 word tokens in the German narratives (out of which 282 words were Swedish) and 6,252 word tokens in the Turkish narratives (out of which 97 words were Swedish). Generally, codeswitches were thus relatively rare. Note that the number of words cannot straightforwardly be compared across languages, as what constitutes a word varies greatly between the languages. In the case of the three languages studied here, one main difference is the agglutinating nature of Turkish, where a proposition which requires several words in e.g. Swedish or German can be expressed using only one Turkish word. To give another example of differences between the languages, Swedish has definite suffixes whereas German has definite articles; this means that every definite NP consists of two words in German, but only one in Swedish. The lower number of words in Turkish, and the higher number of words in German, compared with the same children's Swedish production should thus not be taken to mean that the Swedish-Turkish bilinguals are less productive in Turkish or that the Swedish-

German are more productive in German. Instead it is an expected result based on the differences between the languages.

4.3 Words per narrative

The number of word tokens per narrative is included as a variable in the analyses carried out in Chapter 8, and for this reason, results for this variable are reported here. Table 4.5 shows the number of words per narrative in the Swedish narratives, by language and age group.[53]

Table 4.5. Words per narrative, Swedish narratives, by age group, language group and narrative task.

		MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)
Monolinguals (N=72)				
4-year-olds	Mean (SD)	72.3 (29.3)	60.3 (21.9)	66.1 (25.8)
	Range	34 – 135	23 – 118	28 – 134
5-year-olds	Mean (SD)	74.3 (19.6)	64.0 (22.2)	82.4 (26.6)
	Range	45 – 123	39 – 112	46 – 161
6-year-olds	Mean (SD)	89.9 (24.9)	78.9 (23.2)	94.0 (31.5)
	Range	56 – 146	36 – 132	45 – 160
Swedish-German bilinguals (N=46)				
4-year-olds	Mean (SD)	68.6 (22.4)	56.9 (15.6)	59.5 (18.1)
	Range	26 – 124	38 – 84	29 – 86
5-year-olds	Mean (SD)	66.1 (24.3)	66.0 (26.5)	71.4 (22.1)
	Range	27 – 114	26 – 110	45 – 116
6-year-olds	Mean (SD)	91.8 (24.8)	81.3 (29.3)	91.0 (39.9)
	Range	54 – 138	45 – 151	54 – 186
Swedish-Turkish bilinguals (N=48)				
4-year-olds	Mean (SD)	69.9 (45.7)	69.6 (30.0)	86.3 (41.5)
	Range	19 – 211	36 – 164	33 – 166
5-year-olds	Mean (SD)	82.4 (32.5)	89.4 (48.5)	102.2 (44.6)
	Min – Max	18 – 139	36 – 227	43 – 228
6-year-olds	Mean (SD)	75.0 (25.2)	82.3 (30.7)	93.1 (51.5)
	Range	44 – 134	49 – 159	41 – 250

Note. BB = Baby Birds, BG = Baby Goats.

In order to determine if there are differences between the language and age groups in narrative length (words per narrative), three factorial ANOVAs were carried out, one for each narrative task. For MAIN1, there was a signif-

[53] The analysis of words per narrative was carried out on the total number of words in the target language, i.e. for calculations of words per narrative in the German narratives, only German words are included (cf. Table 4.3).

icant difference between the age groups ($F(2, 156) = 4.230$, $p = .016$, $\eta_p^2 = .051$), but not between the language groups ($F(2, 156) = .264$, $p = .769$, $\eta_p^2 = .003$). The post-hoc test for age group showed that the four-year-olds produced shorter MAIN1 narratives than the six-year-olds, but that there were no other group differences. Also in MAIN2 and ENNI did the six-year-olds produce longer narratives than the four-year-olds (MAIN2: $F(2, 156) = 4.533$, $p = .012$, $\eta_p^2 = .055$; ENNI: $F(2, 156) = 5.292$, $p = .006$, $\eta_p^2 = .065$). In both MAIN2 and ENNI, there were significant effects of language group (MAIN2: ($F(2, 156) = 30.826$, $p < .001$, $\eta_p^2 = .282$; ENNI: $F(2, 156) = 3.879$, $p = .023$, $\eta_p^2 = .049$), but the post-hoc tests showed that the effect was not identical for the two tasks. The Swedish-Turkish bilinguals produced significantly longer MAIN2 narratives than both other groups, and the monolinguals produced longer MAIN2 narratives than the Swedish-German bilinguals. For ENNI, the only significant difference was between the two bilingual groups; the Swedish-Turkish bilinguals produced longer ENNI narratives.[54] There were no interaction effects between age and language group.[55] In all groups and for all narrative tasks, there is substantial individual variation in narrative length.

Table 4.6 shows the number of word per narrative in the German data, for the three age groups of the Swedish-German bilinguals. Running one-way ANOVAs on the length of the narratives in words showed no significant differences between the age groups for any of the narrative tasks (MAIN1: $F(2, 43) = 1.322$, $p = .277$; MAIN2: $F(2, 43) = 2.676$, $p = .08$; ENNI: $F(2, 41) = 1.534$, $p = .228$). In all groups, and for all three tasks, there was large variation between individual children.

Table 4.6. Words per narrative, German narratives, Swedish-German bilinguals (N=46), by age group and narrative task.

		MAIN1 (Cat/Dog)	**MAIN2 (BB/BG)**	**ENNI (A2/B2)**
4-year-olds	Mean (SD)	95.5 (44.4)	72.9 (33.2)	84.3 (27.2)
	Range	35 – 188	41 – 169	39 – 139
5-year-olds	Mean (SD)	85.0 (52.2)	62.3 (22.5)	90.6 (47.8)
	Range	31 – 194	30 – 99	44 – 203
6-year-olds	Mean (SD)	112.7 (48.3)	88.8 (39.6)	114.4 (62.2)
	Range	53 – 213	42 – 201	42 – 303

Note. BB = Baby Birds, BG = Baby Goats.

In Table 4.7, results for words per narrative in the Turkish narratives are shown. One way ANOVAs showed no significant differences in length between the age groups, neither for MAIN1 ($F(2, 45) = .773$, $p = .468$) nor for

[54] P-values for all pairwise comparisons are reported in Appendix 2, Tables A2.5-A2.6.
[55] MAIN1: $F(2, 156) = 1.470$, $p = .214$, $\eta_p^2 = .036$; MAIN2: $F(2, 156) = 1.018$, $p = .400$, $\eta_p^2 = .025$; ENNI: $F(2, 156) = .920$, $p = .454$, $\eta_p^2 = .024$.

MAIN2 ($F(2, 45) = .723$, $p = .491$). In Turkish, just as in German, variation between individual children was large.

Table 4.7. Words per narrative, Turkish narratives, Swedish-Turkish bilinguals (N=48), by age group and narrative task.

		MAIN1 (Cat/Dog)	**MAIN2 (BB/BG)**
4-year-olds	Mean (SD)	57.2 (24.7)	62.1 (29.9)
	Range	33 – 107	26 – 133
5-year-olds	Mean (SD)	59.6 (16.1)	73.5 (28.5)
	Range	13 – 80	14 – 130
6-year-olds	Mean (SD)	65.6 (17.4)	66.6 (21.7)
	Range	39 – 87	39 – 126

Note. BB = Baby Birds, BG = Baby Goats.

5 Vocabulary

Words are an essential part of a story. Without at least some basic vocabulary it is not possible to narrate any events. Most studies that investigate children's narrative competence therefore include one or several measures of vocabulary. This chapter investigates the children's lexical knowledge in the form of scores on the production parts of the Cross-linguistic Lexical tasks (CLTs) in Swedish, German and Turkish, and the children's narrative vocabulary in the MAIN narratives. The aim is to gain a better understanding of effects of age and, for Swedish, language group on vocabulary production scores and narrative vocabulary and to investigate the relationship between the two types of lexical measures. The relationship between vocabulary and macrostructure is analyzed in Chapter 8.

This chapter proceeds as follows. After a summary of insights from earlier studies of vocabulary in bilingual preschool children (Section 5.1), results for vocabulary production (Section 5.2) and narrative vocabulary (Section 5.3) are presented. The chapter closes with a discussion of the findings (Section 5.4).

5.1 Vocabulary in bilingual preschool children

In this section, central findings from earlier studies of vocabulary in bilingual children are summarized, focusing on studies of the development of vocabulary production (expressive vocabulary) in the late preschool and early school years. Central aspects that explain results of different groups beside age are socio-economic status (SES) and amount of input. Both aspects have been investigated in a number of studies and are generally thought to be important for explaining bilinguals' vocabularies in the two languages. First a general summary is given, after which the section is further divided into different subsections, each dealing with studies whose results are central to the current study. The first subsection deals with studies using Cross-linguistic lexical tasks (CLTs) (Section 5.1.1), the same vocabulary task as in the current study. The second and third subsections describe studies of the populations included in the current study, German- and Turkish-speaking bilinguals (Section 5.1.2), and mono- and bilingual Swedish-speaking children (Section 5.1.3). Finally, the last subsection summarizes studies of vocabulary in narratives (Section 5.1.4).

Many studies have investigated vocabulary in bilinguals aged 1–3 using MacArthur-Bates Communicative Development Inventories (CDIs, Dale, 2007; Fenson et al., 1993), which are vocabulary checklists where parents mark the lexical items their child comprehends and produces. These studies concern very young bilinguals and a very different method of data collection and their results are therefore not central to the current study. However, the results do point to the importance of input for vocabulary development in both languages and indicate that vocabulary in the majority language develops faster than in the minority language as bilinguals grow older (Hoff et al., 2012; Hoff, Rumiche, Burridge, Ribot, & Welsh, 2014; Pearson, Fernandez, Lewedeg, & Oller, 1997).

In studies of older bilingual children, just as in the current study, vocabulary tests are often used. Earlier studies have compared bilinguals with monolinguals, and analyzed effects of SES and amount of input in the two languages, as well as development with age. Studies of vocabulary comprehension appear to be more common than studies of vocabulary production. Comprehension and production are rarely investigated using one single test; the Cross-linguistic Lexical Task (CLT) is an exception.

The most commonly used vocabulary test is the Peabody Picture Vocabulary Test, the PPVT (Dunn & Dunn, 1997) testing vocabulary comprehension. The PPVT and its adaptations to different languages have been used in a number of studies of bilinguals (e.g. Bialystok, Luk, Peets, & Yang, 2010; Buac et al., 2014; Calvo & Bialystok, 2014; Cobo-Lewis, Pearson, Eilers, & Umbel, 2002a, 2002b; Dijkstra, Kuiken, Jorna, & Klinkenberg, 2016; Gathercole & Thomas, 2009; Gathercole, Thomas, Roberts, Hughes, & Hughes, 2013; Thordardottir, 2011).

Results from vocabulary comprehension in the majority language are mixed. Some studies point to differences between mono- and bilinguals across age groups (e.g. Bialystok et al., 2010), and indicate that some bilingual groups may be as much as three years behind their monolingual peers (Vermeer, 2001). Other studies found no differences between monolinguals and bilinguals in vocabulary comprehension when bilinguals had at least 40% exposure to the language (Thordardottir, 2011). Additionally, Socioeconomic status (SES) also influences vocabulary; mid-SES bilinguals tend to perform similarly to monolinguals from low-SES backgrounds (Calvo & Bialystok, 2014). Both SES and amount of input are thus important factors to take into consideration when comparing groups of bilingual children.[56] With regard to age development in vocabulary comprehension, there are indications that gains in the minority language are smaller than in the majority language (Gathercole et al., 2013).

[56] The importance of SES and input and especially the link between them, with higher SES meaning more (quality) input which leads to larger vocabulary in children, has also been shown for monolinguals, e.g. Hart & Risley (1995).

Although studies of vocabulary production in bilinguals are less common, results point to the importance of both input and SES. For example, in a study of 90 Russian-German bilinguals growing up in Germany and 79 Russian-Hebrew bilinguals growing up in Israel, Gagarina et al. (2014) analyzed effects of age, length of exposure and input on verb and noun production. The children were 4 to 6 years old and most had two L1 Russian parents. Age correlated significantly with scores in the minority language Russian in both groups, but the increase with age was relatively small (cf. Gagarina et al., 2014, Fig. 4.3a, p. 74). There was a steep increase with age in German for the Russian-German group, but no increase with age in Hebrew for the Russian-Hebrew group, which was likely linked to the children's already high performance in Hebrew. The Russian-Hebrew children performed better than the Russian-German bilinguals in both languages, which may have been due to the fact that they had higher SES. The results from this study point to differences between different bilingual groups as well as differences in age development in the majority and the minority language. Dijkstra, Kuiken, Jorna, & Klinkenberg (2016), in their longitudinal study of bilinguals learning Dutch and Frisian, two closely related languages with a high number of cognates, and growing up in the bilingual province Friesland in the Netherlands, found similar results. In this study, three rounds of testings were carried out with intervals of approximately six months with the first testing at age 2;6–2;11 and in both languages, the children scored higher as they got older. However, age development was not the same in the two languages. Children who spoke mainly Frisian at home were developing expressive Frisian vocabulary at a faster rate than children with Dutch at home, but the increase in the minority language Frisian was generally slower compared with the majority language Dutch. It thus seems that already at this young age, there is an effect of majority language influence, even for children exclusively speaking a closely-related minority language with relatively high status at home. Similar results were found in a longitudinal study of Spanish-English bilinguals growing up in the U.S., where the children's expressive vocabulary in English developed between age 5 and 6, but their Spanish did not (Uccelli & Páez, 2007). In a recent study, Haman, Wodniecka, et al. (2017) found that Polish-English bilinguals growing up in the U.K. developed their productive vocabulary in the minority language at a slower pace compared with SES-matched Polish monolinguals in Poland, and also that bilinguals who received more Polish input developed their Polish vocabulary at a faster pace than those who received less.

In a study comparing Russian-Hebrew bilingual and Hebrew monolingual 5–6-year-olds growing up in Israel, Meir & Armon-Lotem (2017) found significant effects of both SES (low vs mid-high) and bilingualism on vocabulary production in the majority language Hebrew, with children from low-SES backgrounds performing lower than those from middle-high-SES backgrounds and bilinguals performing lower than monolinguals. The effect

of bilingualism was stronger than the effect of SES. There was no effect of SES on vocabulary production in the minority language Russian.

To summarize, earlier studies have shown effects of SES and language input on scores on bilingual children's vocabulary in both languages. Higher SES is often linked to more input in the majority language, which leads to better results on vocabulary tests in that language. In many studies, bilinguals performed lower than monolinguals (in some cases with similar SES-backgrounds) in one or both languages (Bialystok et al., 2010; Buac et al., 2014; Haman, Wodniecka, et al., 2017; Meir & Armon-Lotem, 2017; Vermeer, 2001; see also the studies on Turkish-Dutch bilinguals below). More input in one language leads to better scores in that language (Cobo-Lewis et al., 2002a, 2002b; Dijkstra et al., 2016; Haman, Wodniecka, et al., 2017). Results indicate that the effect of age is different for the two languages. Whereas many studies show a stable age development in the majority language (e.g. Bialystok et al., 2010; Cobo-Lewis et al., 2002a; Dijkstra et al., 2016; Uccelli & Páez, 2007), children seem to make smaller gains in the minority language, especially for language production, even when they receive substantial input in that language and/or the minority language has relatively high status in the surrounding society (Dijkstra et al., 2016; Gagarina et al., 2014; Gathercole et al., 2013; Hoff et al., 2012).

5.1.1 Studies using Cross-linguistic Lexical Tasks (CLTs)

Although the CLTs (for a description of the test and its procedure, see Section 3.2.1) are relatively new, a number of studies have used it to investigate mono-, bi- and even trilingual children, including two studies that compare typically-developing children with children with SLI (Kapalková & Slančová, 2017; Khoury Aouad Saliby et al., 2017).

In a large-scale study with the purpose of testing the usefulness of the CLTs, Haman, Łuniewska, et al. (2017), compared results from 17 different languages (Afrikaans, Catalan, British English, South African English, Finnish, German, Hebrew, isiXhosa, Italian, Lithuanian, Luxembourgish, Norwegian, Polish, Serbian, Slovak, Swedish, Turkish). Data from 639 monolingual children aged 3;0–6;11 were included, ranging from 10 to 89 children per language.[57] Most of the children had mid- to high-SES backgrounds. The results from isiXhosa (which were only available for three CLT-parts and 10 participants) were considerably lower than for all other languages. Differences in scores between the other 16 languages were small. Comprehension scores were higher than production scores for all 16 languages. Sig-

[57] Of the 32 Swedish participants in Haman, Łuniewska, et al. (2017), 26 (24 five-year-olds, two four-year-olds) were the same as in the current study.

nificant age effects were found for 11 languages, including Swedish.[58] For Turkish (N=33), age only correlated significantly with verb comprehension. There was no significant correlation between age and CLT scores for German. The lack of correlation in German is explained by the relatively small sample size (N=36) and the narrow age range (5;0–6;3).

Altman, Goldstein & Armon-Lotem (2017) found clear differences between monolingual Hebrew (N=26) and Hebrew-English (N=27) bilingual children aged 4;9–6;6 growing up in Israel for both vocabulary comprehension and production on the Hebrew CLT; the monolinguals performed better than the bilinguals. Unfortunately, age effects were not investigated in this study, leaving the question open as to how bilinguals' vocabularies develop in relationship with those of monolinguals.

Bohnacker et al. (2016) analyzed effects of age, SES and minority language input on CLT production scores in the minority language of 40 Swedish-Turkish bilinguals (aged 4;0–6;9) and 38 Swedish-German bilinguals (aged 4;0–6;11) growing up in Sweden.[59] Children whose parents spoke mainly or only the minority language both to each other and to the child scored significantly better than children whose parents did not. For children who received less minority language input at home, having minority-language-speaking friends boosted their minority language vocabulary. SES, which could only be tested in the Swedish-Turkish group,[60] did not influence the minority language scores. There was no significant difference between the scores in Turkish and German, which was surprising, given that most Swedish-Turkish children had two parents who were Turkish speakers and this would have been expected to boost the scores in the minority languages. The authors suggest that the high number of cognates between Swedish and German may have a positive impact on the children's German vocabulary; knowing the Swedish word would in many cases also help the child in German. In contrast, the Swedish-Turkish children would not be helped on the Turkish task by their knowledge of Swedish. These results indicate that language distance is a factor that should be taken into account when comparing different groups of bilingual children.

In a study of Maltese-English bilingual five-year-olds, Gatt, Attard, Łuniewska, & Haman (2017) found effects of language dominance on CLT scores in both languages, with children who were English-dominant scoring higher on the CLT in that languages and vice versa, indicating that language dominance is a factor that needs to be taken into account when interpreting

[58] Clear age effects were also found for both comprehension and production in a study of the same 72 Swedish monolinguals as in the current study (Lindgren, 2017).
[59] All these Swedish-German bilinguals and 30 of the Swedish-Turkish bilinguals (ten in each age group) were the same as in the current study.
[60] There was no variation in SES in the Swedish-German group; all children in this group came from high-SES backgrounds (see Section 3.1.2.1 for details).

bilinguals' scores in their two languages, irrespective of the performance of monolingual groups in different languages.

To summarize, results from earlier studies using CLTs show clear age effects for monolinguals (Haman, Łuniewska, et al., 2017). Bilinguals may not perform as well as monolinguals in the majority language (Altman et al., 2017). For bilinguals, there may be no clear development with age in the minority language (Bohnacker et al., 2016); age effects in the majority language have so far not been investigated. Similarly to what has been found in studies using other types of vocabulary tests, bi- and trilinguals' vocabulary scores are influenced by the amount of input (Bohnacker et al., 2016; Gatt et al., 2017; Potgieter & Southwood, 2016), but there may also be effects of language distance, with children speaking closely-related languages having an advantage in the minority or less dominant language (Bohnacker et al., 2016).

5.1.2 Bilingual children speaking German or Turkish

With the exception of the study by Bohnacker et al. (2016) described above, there are no earlier studies of vocabulary in Swedish-Turkish or Swedish-German children. A number of studies have been carried out on other bilingual children speaking Turkish or German. Studies of German-speaking bilingual children have primarily had German as the societal language, such as the Russian-German bilinguals of Gagarina et al. (2014). What to expect of children speaking German as a minority language is unknown. Concerning Turkish-speaking bilinguals, most studies have been carried out on children belonging to the Turkish-speaking diaspora, mainly in Germany and the Netherlands.

Several studies have analyzed vocabulary development in Turkish-Dutch bilingual children growing up in the Netherlands (e.g. Leseman, 2000; Prevoo et al., 2014; Scheele, Leseman, & Mayo, 2010), although they rarely measure vocabulary production in both languages. For example, in the study of Scheele et al. (2010), only comprehension was analyzed. This study of children aged 2;11–3;7 found a large difference between Dutch-Turkish bilinguals and Dutch monolinguals. The monolinguals scored better in Dutch than the bilinguals in Turkish. Although the age range was relatively large, age effects were unfortunately not analyzed. Prevoo et al. (2014) analyzed vocabulary production in Dutch and vocabulary comprehension in Turkish of 111 Turkish-Dutch children aged 5;5–6;10. They found that children with higher SES and those with more Dutch input had higher scores on Dutch vocabulary production. There was no effect of SES on Turkish vocabulary comprehension, but a clear effect of amount of Turkish input. Importantly,

age correlated with the children's scores in the majority language Dutch but not in the minority language Turkish.[61]

A study that did investigate both comprehension and production using parallel tests in Turkish and Dutch is Leseman (2000). In this longitudinal study of 31 low-SES Turkish-Dutch children and 77 Dutch monolinguals (31 low-SES, 46 high-SES), the children were tested three times with approximately six months in between from age 3;2 to 4;2. At the first testing, expressive vocabulary could not be tested due to the bilinguals' limited knowledge of Dutch and even their scores on Dutch vocabulary comprehension were low. There was a clear increase in the bilinguals' Dutch vocabulary with age, but they had a slower development and performed much lower than the monolingual children with similar low-SES backgrounds at age 4.[62] The bilinguals' scores in Turkish were similar to the scores of the Dutch monolinguals with low-SES backgrounds. The amount of Dutch input significantly affected the children's Dutch scores.

Akoğlu & Yağmur (2016) compared the Turkish of 30 Turkish-Dutch bilingual 5-year-olds growing up in the Netherlands, to 30 age- and gender-matched monolinguals with comparable social backgrounds in Turkey. They analyzed a number of different linguistic measures, including vocabulary comprehension and production. The differences in vocabulary scores of the two groups were relatively large, and larger for production than for comprehension. SES as measured by maternal education influenced vocabulary production in the bilinguals; the higher the mother's education, the better vocabulary production score.

Studies on Turkish-speaking bilinguals thus point in the same direction as other studies of bilingual children, namely that amount of input plays an important role for bilinguals' knowledge of both languages, but also that age and SES influence vocabulary production scores, though possibly less so for the minority language than for the majority language. All these aspects need to be kept in mind when interpreting the results from the current study.

5.1.3 Studies of mono- and bilingual Swedish-speaking children

There are few studies of vocabulary in Swedish-speaking children. Some studies of young monolingual children have been carried out using Swedish CDIs, i.e. parental checklists (Berglund & Eriksson, 2000; Cox Eriksson, 2014; Eriksson, 2017). Some studies of older (school-aged) children have investigated the link between vocabulary and reading comprehension (e.g. Lindberg & Johansson Kokkinakis, 2007). A number of unpublished MA

[61] As comprehension was measured in Turkish, but production in Dutch, this could potentially explain the differences with regard to age.

[62] There was a strong effect of SES in the monolinguals, with higher scores for the high-SES group.

theses in speech-and-language pathology have studied preschool and school children's vocabulary (e.g. Brusewitz & Gómez-Ortega, 2005; Krüger Vahlquist, 2012; Mikoczy & Nyman, 2008). A few studies have investigated lexical organization in bilinguals, mainly on word associations using the Kent-Rosanoff list (e.g. Holmström, 2015; Namei, 2002; Salameh, 2011).

For vocabulary production specifically, there are some studies of monolingual Swedish preschool and school children using the Swedish version of the Boston Naming Test, BNT (Tallberg, 2005), a test of production of nouns. For example, Brusewitz & Tallberg (2010) tested 152 Swedish monolingual children evenly divided in Kindergarten (6-year-olds), Grade 3 (9-year-olds), Grade 6 (12-year-olds) and Grade 9 (15-year-olds) with the BNT. They found clear age effects, with significant differences between all the age groups. Age effects on the Swedish BNT scores were also found for 28 children aged 3;2–4;1 (Westlin & Ytterdahl, 2007).

Few studies on Swedish-speaking bilinguals have been published. With the exception of a recent study by Ganuza & Hedman (2017), those that are have not investigated age effects (Holmström, 2015; Salameh, 2011). For example, in her doctoral thesis, Holmström (2015) analyzed receptive and expressive vocabulary in both languages of Swedish-Arabic seven-year-olds with typical development (N=15) and with LI (N=15). Holmström (2015) found differences in vocabulary size between the typically-developing children and the LI group, except for in Arabic vocabulary production. Both groups performed better in Swedish on vocabulary production, but their comprehension scores in the two languages were comparable.

Ganuza & Hedman (2017), investigated effects of chronological age, age of arrival to Sweden and mother-tongue instruction (MTI) attendance on vocabulary and literacy in both languages of 120 Swedish-Somali bilinguals aged 6–12. Vocabulary was investigated by translated versions of the PPVT in both languages and a task containing antonyms, hypernyms, and synonyms aimed at measuring paradigmatic semantic relationships. MTI attendance affected Somali vocabulary and reading positively. For vocabulary, older children scored better in Swedish than in Somali, whereas younger children had similar scores in both languages. On all measures in both languages, older children scored higher than younger children. A subset of the children (N=46) were tested again a year later. Notably, these children scored significantly better at the Swedish, but not at the Somali PPVT. This result indicates a slower development of (receptive) vocabulary in the minority language Somali compared to in the majority language Swedish.

Based on these studies, as well as international ones (e.g. Haman, Łuniewska, et al., 2017), one can expect a clear relationship between vocabulary production scores and age in monolinguals, and possibly for bilinguals in the majority language Swedish. What to expect for the performance in the minority language and regarding the relationship between vocabulary scores in the bilinguals' two languages is less clear.

5.1.4 Studies of vocabulary in narratives

Studies of children's spontaneous speech or narrative tasks often investigate lexical measures (e.g. Iluz-Cohen & Walters, 2012; Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003; Pearson, 2002; Reuterskiöld et al., 2011; Uccelli & Páez, 2007).[63] In addition to productivity measures, such as total number of word tokens, simple measures that seek to capture lexical diversity are often included.[64] Examples of such measures include number of different words (NDW, i.e. number of types) and type-token-ratio (TTR). Few studies have investigated the relationship between narrative vocabulary and macrostructure or compared lexical measures in narratives with data from vocabulary tests. Both these aspects are therefore investigated in the current study.

TTR in particular has been used as a measure of lexical diversity in a number of studies. This measure has been criticized, as it is highly dependent on text length, such that TTR inevitably decreases as texts become longer. For example, Vermeer (2000) tested the efficiency of ten frequently used lexical measures, including number of tokens, number of different words (types), TTR and the Guiraud index ('indice de richesse', Guiraud, 1959), in distinguishing between children with different vocabulary sizes compared to the children's scores on a receptive vocabulary test and a word definition task.[65] Data was collected in Dutch from 70 Dutch L1 and 76 Dutch L2 children (children who all spoke primarily another language than Dutch at home). The children were 4–7 years old and all came from low-SES backgrounds. No correlation was found between TTR in spontaneous speech samples and vocabulary test scores, and TTR could also not distinguish between mono- and bilinguals, even though the monolinguals scored much higher on the vocabulary tests. Measures such as number of types (NDW), tokens and lemmas did show differences both between age groups and between monolinguals and bilinguals. Vermeer (2000) argued that, for this reason, NDW is a good measure for distinguishing between children with different lexical proficiency (cf. Uccelli & Páez, 2007). NDW was used as a simple measure of lexical diversity in the current study.

Two studies of English-Spanish bilinguals growing up in the U.S. using NDW as a measure of lexical diversity have found partly conflicting results with respect to age development. In Frog story narratives collected in English, Muñoz et al. (2003) found no differences in NDW between four- and

[63] Lexical measures also form a part of some assessment protocols for narratives, such as the Index of Narrative Microstructure (Justice et al., 2006).

[64] A number of more specific measures have also been investigated, such as story-specific nouns (Pearson, 2002), verb use (Viberg, 2001, 2004), number of different verbs per c-unit (Reuterskiöld, Hansson, & Sahlén, 2011), and the number of adverbs and elaborated noun phrases (Greenhalgh & Strong, 2001). In the current study, no such specific measures are analyzed, and therefore results from this type of study are not described here.

[65] For another assessment of the usefulness of different lexical diversity measures, see McCarthy (2005).

five-year-olds. In contrast, Uccelli & Páez (2007) found a difference between age five and six for the majority language English, but not for the minority language Spanish. It should thus be kept in mind with regard to the current study that age development in narrative vocabulary may not be linear, and that vocabulary development in the two languages can be different.

5.2 Vocabulary Production (CLTs)

In this section, results for vocabulary production, i.e. scores from the CLTs, are reported. The reason why only CLT production scores (noun production and verb production) are reported is that vocabulary production can be assumed to be more closely linked to narrative production than vocabulary comprehension.[66] After a description of coding and analyses (Section 5.2.1), results for the Swedish vocabulary production for the different language and age groups are compared (Section 5.2.2). The next sections contain results for both languages of the Swedish-German bilinguals (Section 5.2.3) and the Swedish-Turkish bilinguals (Section 5.2.4). Finally, a comparison of the results from Turkish and German is carried out (Section 5.2.5), followed by a summary of the results (Section 5.2.6). The following research questions are asked:

- Are there differences between age groups and, for Swedish, between language groups on vocabulary production scores from the CLTs?
- Is there any difference between the bilinguals' vocabulary production scores in the two languages?
- Is there any difference between the vocabulary production scores in the minority languages German and Turkish?

5.2.1 Scoring and analysis

5.2.1.1 Scoring of the Cross-linguistic Lexical Tasks

For each language version of the CLT (Swedish, German and Turkish), scoring was done by linguists and/or SLPs who were native or near-native speaker of the language (in addition to the author, the scorers were Ute Bohnacker, Sibylle Dillström and Buket Öztekin). All scorers had extensive knowledge of child language development and testing and had been experimenters in the current study.

The maximum score on the CLT production is 60 points (30 points on noun production and 30 points on verb production). One point was awarded

[66] For a study of both comprehension and production results from the CLT for the monolingual children, see Lindgren (2017), and for the Swedish-German bilinguals, see Lindgren & Bohnacker (submitted).

for each correct answer. In addition to producing the target word, the following cases were also scored as correct: (1) adult-like synonyms (e.g. Ger. *fegen*/*kehren* 'to sweep', Tur. *cizme*/*bot* 'boot', Ger. *Tor*/*Pforte* 'gate'), (2) words that corresponded to the picture but were more specific than the target word (e.g. Swe. *meta* 'to angle' for the target *fiska* 'to fish', Swe. *champinjon* 'button mushroom' for the target *svamp* 'mushroom', Tur. *yemek pişiriyor* 'food cooking' for the target *yemek yapıyor* 'food making', Ger. *Uhu* 'eagle-owl' for target *Eule* 'owl'), and (3) word forms with a pronunciation that was slightly off-target when compared to the (adult) norm either because of a foreign accent (e.g. a child producing *motar* for Swe. *matar* 'to feed', or *Pingvin* for Ger. *Pinguin*) or forms with typical child language simplification errors such as dentalization (e.g. child produces [du:de] for Ger. *Gurke* 'cucumber', simplifying the syllable structure and replacing the velar consonants by dental ones). All other types of responses were scored as incorrect. These included forms that were so different from the adult norm that they were not recognizable to a native-speaker listener with knowledge of child language development of that specific language, e.g. because they were too phonologically and/or morphologically deviant from the target word (e.g. *şengürü* for Tur. *kanguru*, corresponding to the Swedish word *känguru*, but unrecognizable to a Turkish native speaker with no knowledge of Swedish). Other types of answers scored as incorrect included producing a more general word than the target item (e.g. *jobba* 'work' for Swe. *bygga* 'build', *fågel* 'bird' for Swe. *uggla* 'owl') as well as paraphrases and explanations.

Scoring focused on production of the correct lemma. Whether or not the child produced an indefinite, definite or a bare form (i.e. *en banan* 'a banana', *bananen* 'the banana', or *banan* 'banana') and whether or not a noun had the correct gender (i.e. *en banan* or **ett banan* 'a banana') was not taken into account. Similarly, for verbs, tense did not influence the scoring (e.g. *sitta* 'to sit' or *sitter* 'sit(s)'), as long as a form recognizable as the correct lemma was produced. If a child produced a (morphological/phonological) form belonging to a different lemma (e.g. present tense *sticker* meaning 'sting(s)' for Swe. *stickar* 'knit(s)'), it was scored as incorrect. If a child produced a word from a different part of speech (e.g. *snö* 'snow' for Swe. *snöa* 'to snow', *kayıyor* 'sliding' for Tur. *kaydırak* 'a slide'), it was scored as incorrect.

For most child responses, scoring was simple and straightforward. In the cases when it was not, an extensive procedure of discussion and rechecking was applied, to achieve consistency across languages and raters. Whenever deemed necessary, the recordings of the child's responses were consulted. This included cases where the child's answer had been marked as unclear by the experimenter on the test form or where the answer written down could not be identified as a word in the target language or another language known to the members of the research group. If a bilingual child had many answers in the other language, all the child's answers were double-checked against

the recordings. Additionally, all answers from a subsample of the participants were listened to again to make sure that the experimenter had written the answers correctly on the test form. All cases for which scoring was not crystal-clear were discussed in the research group. For especially tricky or unclear answers, more than one round of discussions (and listening to the recordings) was carried out. In these cases, the audio recordings were always consulted. For further information, including lists of correct alternative responses in Swedish, German, and Turkish, see the BiLI-TAS project (Bohnacker, 2013) CLT scoring guidelines (Guidelines for scoring CLT, version March 2018).

5.2.1.2 Statistical analyses

All group comparisons were carried out on total production scores only (i.e. scores from the noun and verb production parts combined).

The Swedish scores were first correlated with age (in months) for the language groups separately, after which a factorial ANOVA was run on the scores from all children. The ANOVA included language group and age group as independent variables (factors). For the Swedish-German bilinguals, Swedish scores were compared to their German scores using paired-samples t-tests. The German scores for the three age groups were compared using a one-way ANOVA and the scores were correlated with age (in months). The scores from the bilingual children's two languages were also correlated. Identical analyses were run for the Swedish-Turkish bilinguals. Finally, in the last part of the analysis, the German scores of the Swedish-German children were compared to the Swedish-Turkish children's scores on the Turkish CLT, i.e. a comparison of the two minority languages in the current study. This comparison was done using an independent-samples t-test.

5.2.2 Swedish

Table 5.1 shows descriptive statistics for the vocabulary production scores of the monolinguals, the Swedish-German bilinguals, and the Swedish-Turkish bilinguals.

Table 5.1. Swedish (CLT) vocabulary production scores, by language group.

	Monolinguals (N=72)	**Swedish-German bilinguals (N=46)**	**Swedish-Turkish bilinguals (N=48)**
Mean (SD)	47.2 (6.2)	45.2 (6.3)	31.8 (10.7)
Range	31 – 56	28 – 57	8 – 51

Note. Max = 60 points.

When comparing the scores of the language groups shown in Table 5.1, it becomes clear that the two bilingual groups performed differently in com-

parison with the monolinguals. Albeit somewhat lower, the mean scores of the Swedish-German bilinguals closely mirror those of the monolinguals in terms of spread (e.g. similar SDs and score ranges). The mean scores for both groups are around 75% correct answers. The mean score of the Swedish-Turkish group, on the other hand, is much lower. On average, a Swedish-Turkish child reaches only just above 50% of the maximum. Below it will be tested if these differences are statistically significant. The Swedish-Turkish group differs from the two others also in terms of spread with higher standard deviation, and a larger score range. While the highest-scoring Swedish-Turkish bilingual has a score that is relatively close to the highest scores in the other groups, the lowest score in the Swedish-Turkish group is substantially lower.

For the language groups as wholes, there thus seems to be a large difference between the Swedish monolinguals and the Swedish-German bilinguals, who perform similarly, vis-à-vis the Swedish-Turkish bilinguals, who perform much lower. The question is if this pattern is the same for all age groups, or if it is caused by low-performing children in one Swedish-Turkish age group. Table 5.2 shows the total production results for the three language groups by age group.

Table 5.2. Swedish (CLT) vocabulary production scores, by age and language group.

	4-year-olds	**5-year-olds**	**6-year-olds**
Monolinguals (N=72)			
Mean (SD)	43.6 (5.9)	46.3 (6.3)	51.7 (2.9)
Range	32 – 52	31 – 56	45 – 56
Swedish-German bilinguals (N=46)			
Mean (SD)	42.1 (6.4)	44.2 (5.6)	49.0 (5.3)
Range	28 – 49	28 – 51	35 – 57
Swedish-Turkish bilinguals (N=48)			
Mean (SD)	27.1 (10.9)	30.1 (9.4)	38.3 (8.8)
Range	8 – 47	16 – 44	25 – 51

Note. Max = 60 points.

Although, as pointed out above, the scores of the Swedish-Turkish bilinguals are radically different from those of the other two groups, the effects of age seem relatively similar in the three language groups. First, in all three language groups, the scores of the older groups are higher. This is true for both mean scores and score ranges, with higher lowest and highest scores in the older groups.

Second, standard deviations are smallest in the oldest age groups, suggesting that the scores become more uniform as children grow older.[67] The scores of the monolingual six-year-olds are the most uniform, with small standard deviations and a narrow score range.

Third, for all three groups, the largest difference is found between the five- and the six-year-olds. Although the five-year-olds scored higher than the four-year-olds, the mean differences are small. When comparing scores across age groups, it becomes even clearer how much lower the scores of the Swedish-Turkish bilinguals are: on average, a Swedish-Turkish six-year-old scored five points *below* monolingual children who were two years younger (mean score monolingual four-year-olds = 43.6 vs Swedish-Turkish six-year-olds = 38.3). By contrast, the Swedish-German bilinguals only scored about two points below the monolinguals in all age groups, indicating that their lexical development in Swedish is similar to that of monolinguals.

In addition to analyzing differences between age groups, it is also relevant to look closer at scores of individual children. Figure 5.1 shows the Swedish CLT production scores plotted against age (in months) for children in the three language groups.

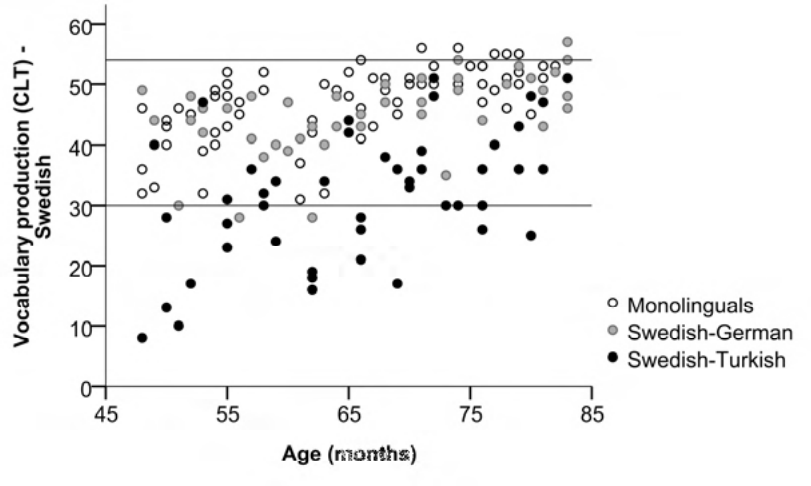

Figure 5.1. Swedish (CLT) vocabulary production scores and the child's age in months. Max = 60 points. A dot may represent more than one individual child. Horizontal lines indicate 50% (30 points) and 90% (54 points).

As Figure 5.1 visualizes, there are positive correlations between age in months and Swedish CLT production scores in all three language groups: generally, older children scored higher (monolinguals: $r = .625$, $p < .001$;

[67] This is also to some extent linked to the fact that the highest scores are close to the test's maximum score.

Swedish-German bilinguals: $r = .492$, $p = .001$; Swedish-Turkish bilinguals: $r = .509$, $p < .001$).

Figure 5.1 also adds further information about the similarities and differences between the groups, in terms of scores of individual children. The monolinguals and Swedish-German bilinguals all performed relatively well on the test. No monolingual child and only two Swedish-German children scored below 50% (30 points). Seven of the older monolinguals (two five-year-olds and five six-year-olds) and three Swedish-German bilinguals (all six-year-olds) scored at or above 90% (54 points or above). With the exception of one Swedish-German child (age: 6;0, score: 35 points), all six-year-olds in these two groups scored at least 43 points, i.e. they knew about 75% of the test items. Only six Swedish-German bilinguals (two four-year-olds, one five-year-old, three six-year-olds) scored outside the monolingual score range for his/her age group. In contrast, 17 of the Swedish-Turkish bilinguals (35% of the children in this group) scored below 50%, and 8 of those even scored at 33% (20 points) or lower. Some of the Swedish-Turkish bilinguals scored as high as the children in the other groups, but many scored much lower. Only two Swedish-Turkish six-year-olds scored below 30 points, compared to eight four-year-olds, and seven five-year-olds.

In order to test the combined effects of the two independent variables on the Swedish CLT production scores, an Age group x Language group (3x3) factorial ANOVA was carried out. There was a significant main effect of language group ($F(2, 157) = 76.059$, $p < .001$, $\eta_p^2 = .492$). The post-hoc tests revealed that, as expected from the descriptive statistics above, there was a difference between the Swedish-Turkish bilinguals and the two other groups.[68] There was also a significant main effect of age group ($F(2, 157) = 22.029$, $p < .001$, $\eta_p^2 = .219$). Here, the post-hoc tests only showed a significant difference between the six-year-olds and the two younger groups.[69] Thus, although the mean for the five-year-olds is somewhat higher than that of the four-year-olds, this difference is too small in relation to the large variation within both these groups to be significant. There was no interaction effect ($F(4, 157) = 0.478$, $p = .752$, $\eta_p^2 = .012$), which means that the differences between the language groups were the same for all age groups and vice versa. There is thus an effect of both age group and language group, and the effect of age is similar in the three language groups. What the multivariate analysis also tells us is that the effect of language group is stronger than the effect of age group (as η_p^2, i.e. the effect size, is larger for language group than for age group). These results can be seen clearly in Figure 5.2.

[68] For p-values of the pairwise comparisons for language group, see Appendix 3, Table A3.1.
[69] For p-values of the pairwise comparisons for age group, see Appendix 3, Table A3.2.

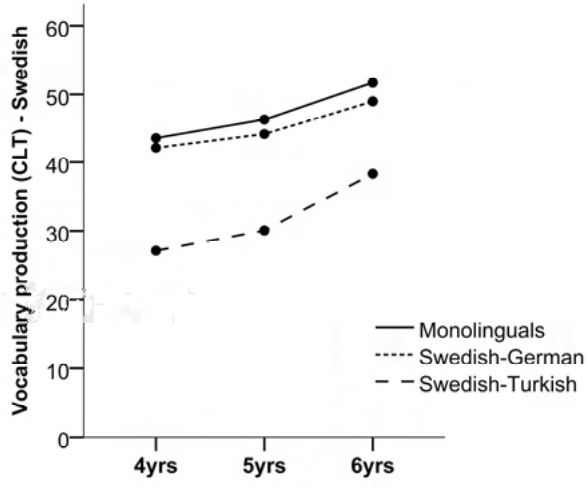

Figure 5.2. Swedish mean vocabulary production scores (CLT), by age and language group.

5.2.3 Swedish-German bilinguals

In Table 5.3, the Swedish-German children's results on vocabulary production for Swedish and German are shown.

Table 5.3. Swedish and German vocabulary production scores (CLTs), all Swedish-German bilinguals (N=46).

	Swedish	**German**
Mean (SD)	45.2 (6.3)	40.1 (11.0)
Range	28 – 57	17 – 56

Note. Max = 60 points.

The Swedish-German bilinguals scored significantly better in Swedish than in German ($t(45) = 2.974$, $p = .005$). On average, the children scored five points higher on the Swedish test. It is striking that the variation in the German scores was larger, in terms of both standard deviations and score ranges. While the highest scores for Swedish and German were comparable, showing that children who performed best in the languages were at an equal level, the lowest score for Swedish is just below 50% (28 points), whereas for German it is just above 25% (17 points). When comparing the scores in both languages for individual children, we see that 17 children scored higher in German, one child had identical scores in the two languages, and the remaining 28 children had higher Swedish scores.

Table 5.4 shows the results on the German CLT for the three age groups. Whereas the Swedish-German bilinguals' scores on the Swedish CLT improved with age (cf. Figure 5.1; Table 5.2), in Table 5.4 we do not see any such clear increase with age for the German scores.

Table 5.4. German vocabulary production scores (CLT), Swedish-German bilinguals (N=46), by age group.

	4-year-olds	5-year-olds	6-year-olds
Mean (SD)	39.1 (9.5)	36.4 (12.6)	44.8 (9.6)
Range	18 – 50	17 – 55	25 – 56

Note. Max = 60 points.

The five-year-olds have the lowest mean scores and the largest variation in the scores. Although there is a relatively large difference in mean scores between the six-year-olds and the younger groups, there is no significant difference between age groups for German vocabulary ($F(2, 44) = 2.545$, $p = .09$). There is also no significant correlation between the German scores and age in months ($r = .257$, $p = .085$), as seen in Figure 5.3.

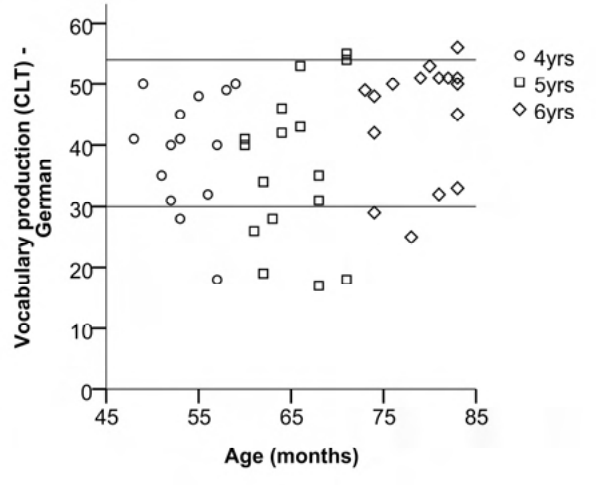

Figure 5.3. German vocabulary production scores (CLT) and the child's age in months. Max = 60 points. Dots may represent more than one individual child. Horizontal lines indicate 50% (30 points) and 90% (54 points).

As can be seen in Figure 5.3, high- and low-performing children were found in all age groups. Most of the children have scores in the 50–90% range. Two four-year-olds, five five-year-olds, and two six-year-olds scored below 50% (30 points). While the scores of the younger children are more evenly spread out, the six-year-old group is different. In fact, the six-year-old group seems to consist of two subgroups, with four children performing just above or just below 30 points (score range 25–33) and the remaining twelve children scoring above 40 points. Most of the children in the higher-performing group are in fact closer to 50 than to 40 points. This indicates that there is a smaller group of older children who have relatively weak German vocabulary, whereas the majority of the six-year-olds have higher scores than the

younger children. In addition to their low scores, these four children[70], according to the parental questionnaires, have similar backgrounds, with relatively little German input,[71] making it even clearer that they form a subgroup.[72] The six-year-olds thus consists of two groups; one which have managed to maintain and develop their German (with high scores both in German and Swedish) and another in which the children's German is less developed, probably due to limited German input in their everyday lives. Removing the scores of the low-scoring subgroup increases the six-year-old mean with a full 5 points to 49.8 and makes the group's scores more uniform (new SD = 3.6 which is only just above a third of the original SD for the six-year-olds, cf. Table 5.4). Rerunning the group comparison without these low-performing six-year-old yields a highly significant difference between the six-year-olds and the younger groups ($F(2, 40) = 6.958$, $p = 0.003$).

Next, the relationship between the scores in the two languages was analyzed. No significant correlation was found between the scores in Swedish and in German ($r = .188$, $p = .210$), indicating that there is no link between the children's knowledge in the one and in the other language. In Figure 5.4, the scores of the children in the two languages are plotted against each other.

In Figure 5.4, there is just one child, a five-year-old, in the bottom-left quadrant, i.e. scoring below 50% in both languages. This child's score was lower in German: 19 points compared with 28 points for Swedish. The child is a trilingual, being exposed to Swedish at daycare and German and Chinese at home, and this could explain why he scored low in both these languages.[73] It is possible that Chinese is his strongest language and that for this specific child, the amount of input he had so far received in Swedish and German was not enough for him to learn more vocabulary. In order to know more about the reasons for his low performance, a detailed case study would be necessary. One child is in the top-left quadrant, scoring slightly above 50% in German and slightly below in Swedish, and one additional child scored exactly 50% in Swedish and somewhat higher in German. Both these children (BiGer4-14, BiGer4-16) are four-year-olds, and considering that, their

[70] BiGer6-01, BiGer6-06, BiGer6-11, BiGer6-18.

[71] The four low-performing children were reported to receive only 20% German input. Their parents spoke mainly Swedish or only some German to each other, and in the case of two children, neither parent used primarily German to speak to them. The children were reported to use very little or no German when speaking to their parents. The parents of three of the children rated the child's German language production to be only 'so-so', whereas the Swedish of all four children was rated as 'very good'. They all scored high on the Swedish CLT.

[72] The same may be said for the four younger children scoring below 20 points (BiGer4-15, 18 points; BiGer5-01, 18 points; BiGer5-04, 19 points; BiGer5-13, 17 points), except that these children scored much closer to their peers; they were not clear outliers within their age groups in the same way as the four six-year-olds.

[73] Note that another trilingual, a six-year-old with the same language combination, scored high in both languages (44 points in Swedish, 50 points in German). The individual variation among trilinguals is likely to be even larger than for bilinguals, especially when only two of the languages are taken into account.

scores are not extremely low in any of the languages. Eight children scored low in German but much higher in Swedish (bottom-right quadrant; all of these children scored at least 40 points in Swedish). Most children (73%, 35 children), however, are found in the top-right quadrant, i.e. they reach at least 50% in both languages. To summarize: most Swedish-German children performed well on vocabulary in both languages, although on average their scores were higher in Swedish than in German. A smaller group (8 children) clearly has stronger Swedish with German lexical knowledge at a relatively low level.

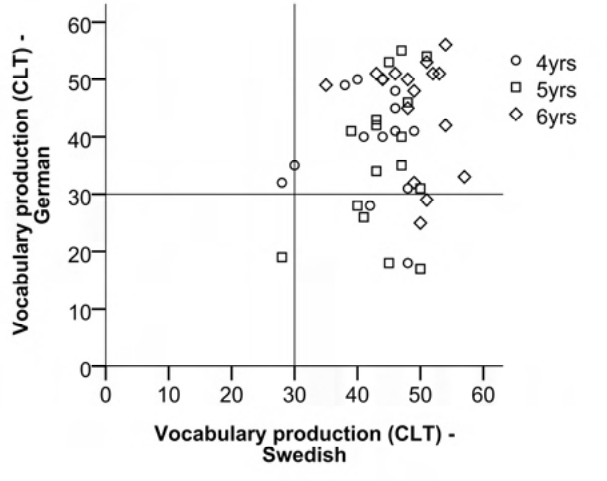

Figure 5.4. Swedish and German vocabulary production scores (CLTs). Max = 60 points. A dot may represent more than one individual child. Lines on both axes indicate 50% (= 30 points).

5.2.4 Swedish-Turkish bilinguals

Table 5.5 shows the Swedish-Turkish bilinguals' results for Swedish and Turkish.

Table 5.5. Swedish and Turkish vocabulary production scores (CLTs), all Swedish-Turkish bilinguals (N=48).

	Swedish	**Turkish**
Mean (SD)	31.8 (10.7)	38.7 (12.3)
Range	8 – 51	3 – 56

Note. Max = 60 points.

The Swedish-Turkish bilinguals scored significantly better in Turkish than in Swedish ($t(47) = -2.572$, $p = .013$). The mean difference between the languages is almost 7 points. Score ranges are wide and SDs are high in both languages, making it evident that there is large variation between individual

children. The Turkish scores have a somewhat larger spread, both measured in SD and score range. In both languages, there are some children who have very low scores as well as some that perform well. However, for the group as a whole, vocabulary is stronger in Turkish than in Swedish. Only 15 children scored better in Swedish, whereas the remaining 33 children had higher scores in Turkish.

Just as for the Swedish-German bilinguals, the Swedish scores of the Swedish-Turkish bilinguals improved with age (see Table 5.2). In Table 5.6, total production scores from the Turkish CLT are shown by age group.

Table 5.6. Turkish vocabulary production scores (CLT), Swedish-Turkish bilinguals (N=48), by age group.

	4-year-olds	**5-year-olds**	**6-year-olds**
Mean (SD)	36.4 (10.9)	37.1 (15.1)	42.4 (10.5)
Range	5 – 52	3 – 56	12 – 53

Note. Max = 60 points.

Interesting to note is that the spread is largest for the five-year-olds, just as was the case for German. For Turkish, both the highest-performing *and* the lowest-performing child is a five-year-old. In the six-year-old group, the lowest score is higher than in the two younger groups, just as was the case for German.

The small differences between the groups shown in Table 5.6, with the older children scoring somewhat higher than the younger, were not statistically significant ($F(2, 46) = 1.113$, $p = .337$). This is probably due to the large spread found in all three age groups. There was no significant correlation with age in months either ($r = .183$, $p = .214$), see Figure 5.5. In Turkish, children with high and low scores can be found in all age groups, similarly to the results for German presented above. Most Swedish-Turkish children (38 children) performed in the 50–90% range, with four additional children having scores just below 50%. Only one child, a five-year-old, performed above 90% (56 points), and two further children, both six-year-olds, scored just below 90% (53 points). Five children, one four-year-old, three five-year-olds and one six-year-old scored below 33% (20 points).[74] It is clear from Figure 5.5 that these children constitute a subgroup with much weaker Turkish than the other children. They all scored at least 1.5 SD below their group means (cf. Table 5.6 above). When looking closer at the backgrounds of these five children, it is not surprising that they scored low in Turkish. Three of the children are trilinguals (including the two children with extremely low scores in Turkish: BiTur4-27: 5 points; BiTur5-21: 3 points), with Kurdish spoken in the family. All had at least 80% Swedish input according to parental report; three even had 95%. Two of the children had parents who grew up in Sweden. All of these children except one (Bi-

[74] BiTur4-27, BiTur5-21, BiTur5-23, BiTur5-24, BiTur6-03.

Tur5-21) scored within the monolingual range on the Swedish CLT. We thus see language imbalance, with Swedish being the stronger language, in three cases combined with trilingualism.

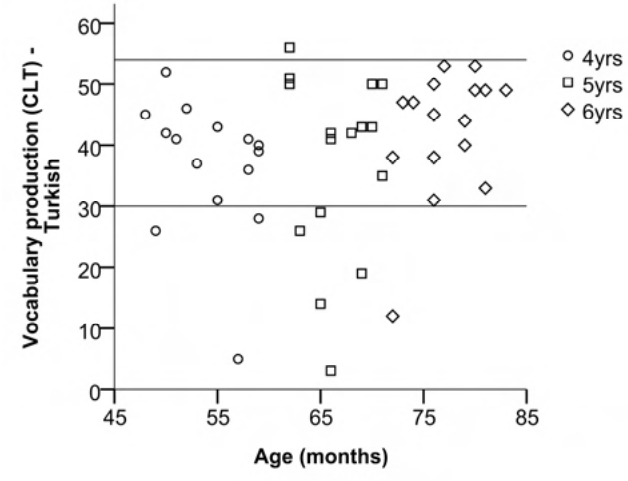

Figure 5.5. Turkish vocabulary production scores (CLT) and the child's age in months. Max = 60 points. Dots may represent more than one individual child. Lines indicate 50% (30 points) and 90% (54 points).

In Figure 5.6, the Swedish-Turkish children's scores in Swedish and Turkish are plotted against each other. The correlation just failed to reach significance (r = -.269, p = .065), but indicates a trend for higher scores in one language meaning lower scores in the other.

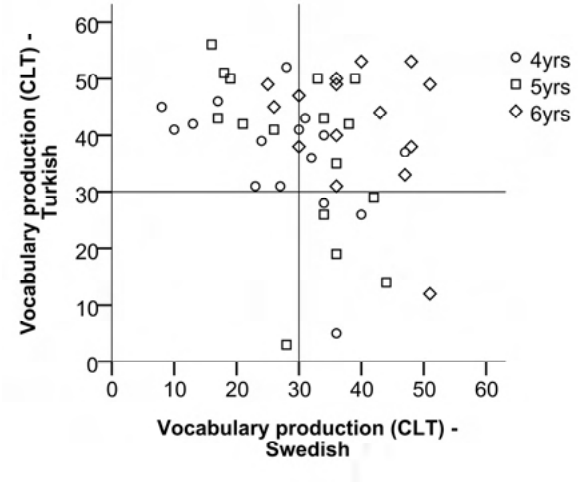

Figure 5.6. Swedish and Turkish vocabulary production scores (CLTs). Max = 60 points. Dots may represent more than one individual child. Lines on both axes indicate 50% (= 30 points).

Figure 5.6 shows that there is only one Swedish-Turkish child, a five-year-old, in the lower-left quadrant, i.e. performing below 50% in both languages. This child (BiTur5-21) scored extremely low (3 points) in Turkish, but has a score of 28 points in Swedish, which is below 50%, but cannot be considered a very low score for a bilingual child in this age group. Eight other children are in the bottom-right quadrant and thus scored below 50% in Turkish, but all of these children have Swedish scores that are comparable to those of monolinguals in their age group. Twenty-two children scored at or above 50% in both languages, whereas 17 children scored above 50% in Turkish but not in Swedish. All except two of the children who do not reach 50% in Swedish are four- and five-year-olds. Thirteen of the six-year-olds perform at or above 50% in both languages.

The picture of the Swedish-Turkish group shown in Figure 5.6 is very different from that shown for the Swedish-German in Figure 5.4. The large difference becomes even clearer when Figures 5.4 and Figure 5.6 are combined, as in Figure 5.7, in which scores in Swedish are plotted against scores in the minority language for the two bilingual groups.

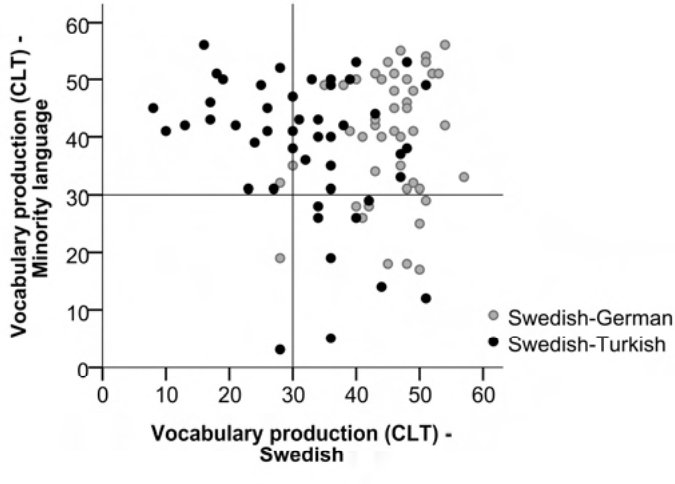

Figure 5.7. Vocabulary production scores (CLTs) for Swedish and for the minority language (Turkish/German), Swedish-German and Swedish-Turkish bilinguals. Max = 60 points. Dots may represent more than one individual child. Lines on both axes indicate 50% (= 30 points).

In Figure 5.7, we see from the relatively small number of dots in the bottom quadrants that most bilinguals scored above 50% in their minority language, German or Turkish. There are similar numbers of children who have Swedish as their stronger language (bottom-right quadrant) in the two bilingual groups. However, the distribution of the groups between the top-right and top-left quadrant is strikingly dissimilar. A substantial number of Swedish-Turkish, but few Swedish-German bilinguals, had relatively weak Swedish

vocabulary production (i.e. scored below 50% on the Swedish CLT production). At group level, the Swedish-German children scored better in Swedish than in German. The Swedish-Turkish children had higher scores in Turkish than in Swedish. The question is whether the Swedish-Turkish children scored better in Turkish than the Swedish-German children did in German. In the next section, scores from Turkish and German are compared.

5.2.5 Comparison of German and Turkish

Table 5.7 shows the results on German and Turkish vocabulary production for the Swedish-German bilinguals and the Swedish-Turkish bilinguals, respectively.

Table 5.7. German and Turkish vocabulary production scores (CLTs), all Swedish-German (N=46) and Swedish-Turkish (N=48) bilinguals.

	German	**Turkish**
Mean (SD)	40.1 (11.0)	38.7 (12.3)
Range	17 – 56	3 – 56

Note. Max = 60 points.

There was no statistically significant difference ($t(92) = .614$, $p = .541$) between the Swedish-German bilinguals' scores in German and the Swedish-Turkish bilinguals' scores in Turkish. Thus, despite the fact that they generally received less minority language input (cf. Section 3.1.2), the Swedish-German bilinguals performed equally well as the Swedish-Turkish bilinguals on vocabulary production in their respective minority language. As mentioned above, it may be the case that the Swedish-German bilinguals were helped by their knowledge of Swedish on the German CLT due to the similarities between the two languages (cf. the discussion in Bohnacker et al., 2016) and therefore performed relatively well on the German CLT.

In Figure 5.8, scores of individual bilingual children in their respective minority language, Turkish or German, are plotted against the child's age in months. Figure 5.8 clearly shows that score patterns in the minority language are similar for the Swedish-German and the Swedish-Turkish bilinguals; in each bilingual group, there is a subgroup of low-performing children and there is no clear age development.

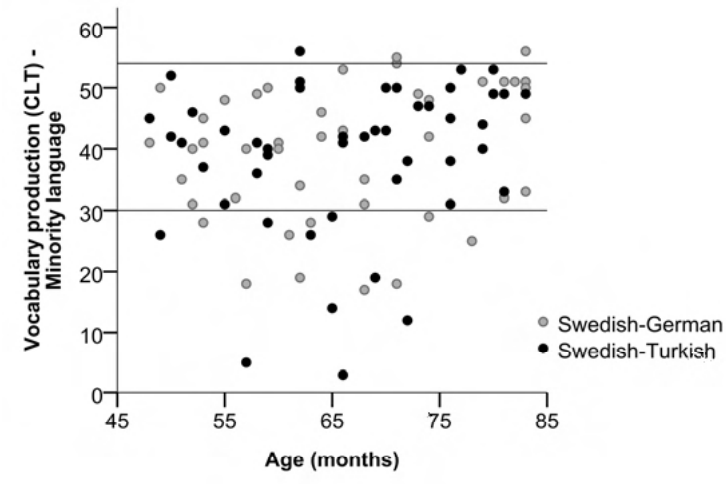

Figure 5.8. Minority language (German/Turkish) vocabulary production scores (CLTs) and the child's age in months. Dots may represent more than one individual child. Lines indicate 50% (30 points) and 90% (54 points).

5.2.6 Vocabulary production: Summary

Beginning with the question of age development, patterns with age were found to be different for the majority language, Swedish, compared with the two minority languages, German and Turkish. Whereas there was a clear development with age in Swedish, this was not the case in the minority languages. For Swedish, there were clear differences in vocabulary production scores between the six-year-olds, who scored better, and the younger groups, who had lower scores. For German, there were only differences between the six-year-olds and the younger children once a subgroup of six-year-olds with especially low scores was removed. In Turkish, there were no effect of age.

For Swedish, there were differences between the language groups: whereas the monolinguals and the Swedish-German bilinguals performed similarly, scores of the Swedish-Turkish bilinguals were substantially lower. In fact, the Swedish-Turkish six-year-olds performed below the monolingual (and Swedish-German) four-year-olds. The development with age was similar in the three language groups. Thus, the Swedish-Turkish bilinguals did not seem to be catching up with their peers in the other language groups.

The Swedish-German bilinguals performed better in Swedish than in German, whereas the Swedish-Turkish bilinguals had higher scores in Turkish than in Swedish. There was no significant correlation between scores in the two languages. The Swedish-German bilinguals performed the same in German than the Swedish-Turkish bilinguals did in Turkish.

5.3 Narrative vocabulary

In this section, the results from the analysis of the children's narrative vocabulary are presented. The number of different words (NDW), i.e. number of word types, in the MAIN narratives was analyzed for all three languages. The aim was to give a broad measure of lexical diversity in the MAIN narratives,[75] to analyze effects of age (and for Swedish, language group), and to relate lexical diversity in the narratives to the vocabulary production scores (on the CLTs). The following two specific research questions are asked:

- Are there any effects of age, and, for Swedish, language group on lexical diversity in the narratives as measured by the number of different words (NDW)?
- What is the relationship between the number of different words (NDW) and vocabulary production scores (CLT)?

The structure of this section is as follows. After describing the analysis carried out (Section 5.3.1), the results are presented for number of different words (NDW) in Swedish (Section 5.3.2), German (Section 5.3.3), and Turkish (Section 5.3.4). The section ends with a summary of the results (Section 5.3.5).

5.3.1 Analysis

The analysis of *number of different words* (NDW) was carried out on the MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) narratives using the frequency (freq) commando in CLAN. Note that 'different word' here does not mean lemma, but that different inflectional forms of the same word count as different words (e.g. *hoppar* 'jump(s)' and *hoppade* 'jumped' count as two different words, although they are forms of the same lemma). Earlier studies have shown there to be very strong correlations between number of different words (word types) and number of lemmas, and it was therefore not judged essential to carry out a lemmatization of all words (cf. Vermeer, 2000, pp. 78–79). All analyses were carried out on NDW for MAIN1 and MAIN2 separately.

For all languages (Swedish, German, Turkish), NDW in MAIN1 and MAIN2 were compared using paired-samples t-tests. For Swedish, in order to investigate effects of language group and age group, two factorial (3x3) Age group x Language group ANOVAs were carried out with MAIN1 and MAIN2 NDWs as dependent variables. For German and Turkish, one-way ANOVAs were carried out to investigate the effect of age group on the

[75] The results for ENNI are not included here as the narrative tasks are different from MAIN (cf. Section 3.2).

MAIN1 and MAIN2 NDWs. For each language, the number of different words in MAIN1 and MAIN2 was also correlated with vocabulary production scores from the CLT. For Swedish, these correlations were carried out for the three language groups separately. No comparisons were carried out between the bilinguals' two languages for this measure as the languages are not comparable in what constitutes a word.

5.3.2 Number of different words: Swedish

In Table 5.8, the results for NDW for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), are shown for the three language groups.[76] For details about each age group within the language groups, see Appendix 3, Table A3.5.

Table 5.8. Number of Different Words (NDW), Swedish: MAIN1 and MAIN2, by language group.

	MAIN1	**MAIN2**
Monolinguals (N=72)		
Mean (SD)	46.1 (12.6)	40.8 (10.5)
Range	24 – 81	20 – 77
Swedish-German bilinguals (N=46)		
Mean (SD)	43.6 (10.6)	40.2 (11.9)
Range	21 – 67	20 – 75
Swedish-Turkish bilinguals (N=48)		
Mean (SD)	38.5 (12.1)	41.8 (12.7)
Range	11 – 67	23 – 76

There was a significant difference in number of different words (NDW) between the two narrative tasks ($t(164) = 2.463$, $p = .015$) for the Swedish narratives. The children produced a higher NDW in MAIN1 (M = 43.5, SD = 12.7), than in MAIN2 (M = 41.2, SD = 11.9). However, the difference was not large (2.3 different words), and the pattern for the Swedish-Turkish bilinguals was the opposite, with a higher NDW in MAIN2.

The factorial (age group x language group) ANOVA for MAIN1 NDW showed a significant effect of both language group ($F(2, 155) = 6.883$, $p = .001$, $\eta_p^2 = .082$) and age group ($F(2, 155) = 9.346$, $p < .001$, $\eta_p^2 = .108$), but there was no significant interaction effect ($F(4, 155) = 2.276$, $p = .064$, $\eta_p^2 = .055$). The effect of age was thus the same in all language groups, and differ-

[76] Swedish MAIN1 data was missing from one Swedish-German five-year-old (cf. Section 4.1). Data from one Swedish-Turkish child (BiTur4-05) was removed from the analysis, as this child's performance was radically dissimilar from that of other children in her group. This was done to not skew the results. This child had an NDW that was 4 and 3 SDs from the group mean for MAIN1 and MAIN2, respectively (MAIN1: 87, MAIN2: 79). The results reported in Tables 5.8 and A3.5 do not include data from this child.

ences between the language groups were found in all age groups. Post-hoc tests showed that there was a significant difference between the monolinguals, who produced a higher NDW and the Swedish-Turkish bilinguals, who produced a lower NDW, but no differences between the Swedish-German bilinguals and the other two groups. There was also a significant difference between the six-year-olds who produced more NDW than both younger groups, who did not differ from each other.[77]

For MAIN2 results were different. The factorial ANOVA showed no significant effect of language group ($F(2, 156) = .303$, $p = .739$, $\eta_p^2 = .004$), but an effect of age group ($F(2, 156) = 13.968$, $p < .001$, $\eta_p^2 = .152$). Post-hoc tests for age group showed clear significant differences between all three age groups.[78] There was no significant interaction effect ($F(4, 156) = .404$, $p = .805$, $\eta_p^2 = .010$); the same age differences were thus present in all three language groups. For number of different words, effects of age and language group are thus different for the two narrative tasks.

Finally, the relationship between Swedish vocabulary production scores (CLT) and NDW from MAIN1 and MAIN2 was analyzed. Here we see interesting differences between the language groups. For the Swedish-Turkish bilinguals, there were significant and strong positive correlations with vocabulary production scores for both MAIN1 ($r = .422$, $p = .003$) and MAIN2 ($r = .440$, $p = .002$) NDW. For the monolinguals, correlations with vocabulary production scores were significant, but differed in strength for MAIN1 ($r = .235$, $p = .047$) and MAIN2 ($r = .347$, $p = .003$), with the correlation for MAIN1 being weaker and just reaching significance. There were no significant correlations between NDW and vocabulary production score in the Swedish-German group, neither for MAIN1 ($r = .247$, $p = .101$) nor for MAIN2 ($r = .283$, $p = .056$). These results indicate that the relationship between expressive lexical knowledge as measured by a vocabulary test and lexical diversity in narratives is not the same for monolinguals and different bilingual groups, at least not in the majority language.[79] The reader should keep in mind that the Swedish-Turkish bilinguals generally scored substantially lower on the CLT than children in the other two groups (cf. Section 5.2.2), with a larger variation in the scores (i.e. a larger score range) and that this may also influence the strength of the correlation between word types and vocabulary scores.

[77] P-values for the pairwise comparisons can be found in Appendix 3, Tables A3.3 and A3.4.
[78] P-values for the pairwise comparison with age are found in Appendix 3, Table A3.4.
[79] Naturally, correlations are also affected by group size; with a larger group of children, the correlation between MAIN1 NDW and CLT production scores would likely have been significant for the Swedish-German bilinguals. Similarly, the same correlation for the monolinguals might not have reached significance if the group size had been smaller.

5.3.3 Number of different words: German

Table 5.9 shows the results for German NDWs, for MAIN1 and MAIN2. For a detailed overview of the results for the different age groups, see Appendix 3, Table A3.6.

Table 5.9. Number of different words (NDW), German MAIN1 and MAIN2, Swedish-German bilinguals (N=46).

	MAIN1	**MAIN2**
Mean (SD)	44.9 (17.1)	38.8 (13.4)
Range	17 – 86	20 – 79

In German, the Swedish-German bilinguals had a significantly higher NDW in MAIN1 than in MAIN2 ($t(45) = 4.054$, $p < .001$). The mean difference was relatively large. For German, the one-way ANOVA showed no differences between the age groups for MAIN1 NDW ($F(2, 44) = 1.939$, $p = .156$) but there was a clear significant difference for MAIN2 NDW ($F(2, 44) = 4.645$, $p = .015$). Post-hoc tests showed that there was a significant difference between the five-year-olds, who produced fewer different words in MAIN2, and the six-year-olds, whereas the four-year-olds did not differ from either of the two older groups.[80] Just as was the case for Swedish, MAIN2 NDW differentiates better between age groups than MAIN1.

There were strong positive correlations with German CLT scores for both MAIN1 NDW ($r = .684$, $p < .001$) and MAIN2 NDW ($r = .566$, $p = < .001$). Children who scored better on German vocabulary production also produced a higher number of different words when telling MAIN narratives in German. For the Swedish-German bilinguals, there is thus a clear link between NDW and CLT scores in the minority language German, but not in the majority language Swedish.

5.3.3 Number of different words: Turkish

Table 5.10 shows the results for NDW in the Turkish narratives for the Swedish-Turkish bilinguals. For results for the different age groups, see Appendix 3, Table A3.7.

Table 5.10. Number of different words (NDW), Turkish MAIN1 and MAIN2, Swedish-Turkish bilinguals (N=48).

	MAIN1 NDW	**MAIN2 NDW**
Mean (SD)	35.4 (10.3)	40.1 (16.3)
Range	9 – 58	10 – 89

[80] For p-values of the pairwise comparison between the age groups, see Appendix 3, Table A3.4.

In Turkish, a paired-samples t-test showed that the Swedish-Turkish bilinguals produced a higher number of different words in MAIN2 than in MAIN1 ($t(47) = -2.274$, $p = .028$). Interestingly, this was the same as their pattern in Swedish.

There was no significant difference between the age groups in Turkish, neither for MAIN1 NDW ($F(2, 46) = .916$, $p = .408$) nor for MAIN2 NDW ($F(2, 46) = .420$, $p = .659$). Age did thus not influence the number of different words used by the child to tell the MAIN narratives in Turkish. Results in all groups are spread out for both narratives.

There was a significant correlation between scores on Turkish CLT vocabulary production and both MAIN1 NDW ($r = .536$, $p < .001$) and MAIN2 NDW ($r = .435$, $p = .002$). Children who scored higher on the Turkish CLT also produced Turkish MAIN narratives which contained a higher number of different words. The children who produced the lowest number of NDW scored among the lowest on the Turkish vocabulary production.

5.3.5 Narrative vocabulary: Summary

First, the effects of age and language group on the number of different words (NDW) depend on which narrative task is used and in which language. In Swedish, MAIN1 NDW showed a significant effect of language group; the Swedish-Turkish bilinguals had a lower NDW than the monolinguals, but there were no differences between the Swedish-German bilinguals and the monolinguals or between the two bilingual groups. The six-year-olds had a higher MAIN1 NDW than the younger groups. For MAIN2 NDW, there were no differences between the language groups, but clear effects of age, with differences between all three age groups: older children produced a higher NDW than younger children. In German, there was an effect of age on MAIN2 only, whereas in Turkish, there were no age effects. These results show that age is a more important factor for performance in the majority language than in the minority language, not only for scores on a lexical task, but also for narrative vocabulary. This will be discussed below together with the differences between the tasks.

Second, the relationship between NDW and vocabulary production scores measured by the CLTs is different for the different groups in Swedish. Whereas there were strong positive correlations between CLT scores and NDW for the Swedish-Turkish bilinguals, the correlation was much weaker for the monolinguals and not significant for the Swedish-German bilinguals. German and Turkish NDWs were highly correlated with the CLT scores in the same languages. In all three languages, the relationships between CLT scores and NDW were the same for both MAIN1 and MAIN2.

5.4 Discussion

This chapter investigated the children's vocabulary. Two vocabulary measures were used, vocabulary production on the Cross-linguistic Lexical Tasks (CLTs) and narrative vocabulary, in the form of the number of different words (NDW), a measure of lexical diversity. In this section, results from each of these measures are discussed in turn.

The first research question for the vocabulary production scores concerned the differences between age groups and, for Swedish, language groups. Starting with the performance of the language groups for Swedish vocabulary production, significant differences were found between the Swedish-Turkish bilinguals, who had a lower performance than the monolinguals and the Swedish-German bilinguals, who did not differ from each other. The Swedish-German children thus performed comparably to the monolinguals on this vocabulary production task. The Swedish-Turkish six-year-olds performed below monolingual and Swedish-German four-year-olds, i.e. they were more than two years behind their peers in the other language groups. This is similar to the results for Dutch-speaking bilingual seven-year-olds who were around three years behind their monolingual peers (Vermeer, 2001). Where do these large differences in majority language vocabulary between the two bilingual groups come from? A number of reasonable (and not mutually exclusive) explanations come to mind.

First, it could be an effect of *SES*. All Swedish-German bilinguals came from a high-SES background, whereas the majority of the Swedish-Turkish bilinguals came from low-SES backgrounds (cf. Section 3.1.2). The monolingual children were also mainly from high-SES backgrounds. SES has been shown to influence vocabulary in the majority language (Calvo & Bialystok, 2014; Meir & Armon-Lotem, 2017). It could thus be the case that the Swedish-Turkish bilinguals' lower scores are a result of their lower SES. Unfortunately, the current study does not allow an investigation of this issue as it does not allow separation of SES and language combination. It was not possible in the current study to match the Swedish-Turkish and Swedish-German bilinguals on SES, as these populations differ in this regard. Future studies of vocabulary in Swedish-speaking children should take SES into account, for example through recruiting a group of low-SES monolinguals for comparison.

A second possible explanation for the differences in performance between the bilingual groups is *language input*. In earlier studies of bilinguals of different ages, the amount of input in the different languages has been shown to be an important factor for vocabulary development (Bohnacker et al., 2016; Gatt et al., 2017; Haman, Wodniecka, et al., 2017; Leseman, 2000; Potgieter & Southwood, 2016). To some extent, this may also be connected to SES, if children with higher SES receive more majority language input. As most children in Sweden (including all children in the current study) attend a

Swedish-medium preschool from an early age, this could level out potential SES-effects on the amount of Swedish input. However, the two bilingual groups in the current study did differ in the amount of daily Swedish input, according to parental report (cf. Table 3.10). There was also a difference between the groups in the age of onset of regular exposure to Swedish with the Swedish-German bilinguals generally receiving regular input in Swedish at an earlier age (cf. Table 3.3). Additionally, in the Swedish-German group, 25 children (out of 46) have one L1-Swedish parent, which is only the case for four (out of 48) Swedish-Turkish children (cf. Table 3.7). More than half of the Swedish-German bilinguals thus receive native Swedish input at home, which likely has given them a distinct advantage in their Swedish lexical acquisition from early on.

A third potentially influencing factor is the *language combination*. Whereas Turkish is typologically distant from Swedish, Swedish and German are closely related languages. Naturally, this influences the number of similarities (e.g. cognates) in the lexicon and could potentially make it easier even for the Swedish-German children who receive less Swedish input to keep up with their monolingual peers (cf. the discussion in Bohnacker et al., 2016). In order to investigate the influence of SES, language input and language combination further, it would be necessary to study other groups, for example of children speaking either typologically close *or* distant languages, but that vary in terms of SES and amount of input. One possibility would be to look at children speaking Swedish and a closely-related language where Swedish is not the majority language, e.g. German-Swedish bilinguals growing up in Germany or Dutch-Swedish bilinguals growing up in the Netherlands. It may also be the case that all these three aspects, SES, input and language combination, together play a role in creating a context for the Swedish-German bilinguals that is boosting their Swedish vocabulary acquisition, and another context for the Swedish-Turkish bilinguals, that is not supporting their Swedish vocabulary acquisition. We will return to the issue of the influence of the language combination when discussing minority language vocabulary below.

The *development with age* in vocabulary production was different for the majority language, Swedish, compared to the two minority languages, German and Turkish. For Swedish, there were clear differences in vocabulary production scores between the six-year-olds, who scored better, and the two younger groups, who had lower scores. For German there were only age differences once a subgroup of older children with low scores were removed, and in Turkish there were no age differences. These results support earlier studies using other types of lexical tasks, both with regard to the clear age effects in the majority language (e.g. Bialystok et al., 2010; Cobo-Lewis et al., 2002a; Dijkstra et al., 2016; Prevoo et al., 2014; Uccelli & Páez, 2007) and less clear development in the minority language (Bohnacker et al., 2016; Dijkstra et al., 2016; Gagarina et al., 2014; Gathercole et al., 2013; Hoff et

al., 2012). It should be pointed out that the scores for 33 Turkish monolinguals aged 4;1–6;10 and 36 German monolinguals aged 5;0–6;3 included in Haman, Łuniewska, et al. (2017) also did not show any significant correlation with age. It is thus possible that the lack of age development in the bilinguals' scores in the minority languages Turkish and German in the current study simply mirrors a similar lack of age development within this age range for monolinguals. However, mean scores for vocabulary production of the Swedish-German and Swedish-Turkish bilinguals were below the marginal mean scores reported in Haman, Łuniewska, et al. (2017) for German and Turkish monolinguals, indicating that the bilinguals of the current study do not perform at the same level as monolinguals in German and Turkish. This issue needs to be investigated further in future studies.

The large differences between the six-year-olds and the younger children found in Swedish require some further comment. Most of the six-year-olds in the current study attended *förskoleklass*, a year that focuses on preparing the children for school, whereas almost all younger children attended preschool. The higher performance of the oldest group may thus be an effect of schooling in addition to age. In order to investigate this, a second study would need to be carried out, in which six-year-olds who have not yet started *förskoleklass* are compared to those who have.

An explanation for the lack of significant differences between the four- and five-year-olds is the high amount of variation between individual children in the younger groups. Although the five-year-olds' mean scores were somewhat higher than the four-year-olds', score ranges were large in both groups. Some four-year-olds performed well, whereas some five-year-olds had relatively low scores. The performance of the six-year-olds showed less spread. Scores on the lexical tasks thus become more uniform as children grow older, something which may partially be linked to ceiling effects.

Two further research questions concerned the relationship between the vocabulary scores in the bilinguals' two languages and whether there were any differences between the scores in German and Turkish. Not surprisingly considering the overall patterns of language proficiency and input in the two bilingual groups (cf. Table 3.4), the Swedish-German bilinguals' performance was better in Swedish than in German, whereas the Swedish-Turkish bilinguals had higher scores in Turkish than in Swedish. This indicates that in terms of lexical proficiency, the Swedish-German bilinguals as a group have Swedish as their stronger language, whereas the Swedish-Turkish bilinguals perform better in Turkish. Interestingly, there was no clear correlation between scores in the two languages. This indicates that vocabulary in the two languages to some extent develops independently, at least at these ages.

Contrary to what one would expect based on input patterns as well as rated language proficiency (cf. Cobo-Lewis et al., 2002a, 2002b; Dijkstra et al., 2016; Haman, Wodniecka, et al., 2017), the Swedish-Turkish bilinguals did

not perform better in Turkish than the Swedish-German bilinguals did in German. Just as for the differences between these groups in the majority language Swedish, a number of possible explanations exist for the lack of differences in the minority languages. These explanations are discussed in Bohnacker et al. (2016), in which results for part of the participants of the current study were presented. Here it is enough to mention that the most likely explanation is linked to the fact that Swedish and German are closely related languages, whereas Turkish is typologically distant. This may impact vocabulary in a number of different ways. First, as Swedish and German share many cognates (i.e. translation equivalents or semantically closely related words that also overlap phonologically), when knowing a word in one language, it may be less demanding to learn that word in the other language. This may thus make vocabulary acquisition in general easier for the child, so that that even with less input in the minority language, it may be possible to acquire the same number of words.

Second, and potentially of equal importance, is the fact that knowing the Swedish word may help the child on the German vocabulary production task. A number of test items, especially on the noun production part, are cognates or even homophonous in Swedish and German. Producing the Swedish word or the Swedish word modified to sound German may help children score points on test items (words) that they actually do not know in German. With Swedish and Turkish, there are no such benefits, no possibilities for 'lucky guesses'. Although care was taken in the current study not to automatically give children a point in German for a Swedish word or vice versa (cf. Section 5.2.1.1), for some items, words in the two languages are more or less homophonous, making it impossible to assess whether the child knows the word in the target language or produced the word in the other language. This means that the Swedish-German children have an advantage on this lexical task and that it may actually over-estimate their German vocabulary.[81] This potential advantage is investigated systematically in a recent study of CLT comprehension and production scores in both languages of the Swedish-German bilinguals (Lindgren & Bohnacker, submitted). The results reported in Lindgren & Bohnacker (submitted) show clear effects of cognate status of the test item in both languages; the Swedish-German children performed better on items that were similar in their two languages (i.e. had a near(-homophonous) translation equivalent in the other language).

The second part of the analysis carried out in this chapter concerned narrative vocabulary. Regarding effects of age and (for Swedish) language group on the number of different words (NDW), the findings show that the effects depend both on the narrative task and on the language. In Swedish, there was no effect of language group on MAIN2 (Baby Birds/Baby Goats)

[81] The same is true for their performance on the Swedish test, where there may be similar benefits from German.

NDW. The Swedish-Turkish bilinguals had a lower MAIN1 (Cat/Dog) NDW than the monolinguals, but there were no other differences between the bilingual groups or between the Swedish-German bilinguals and the monolinguals. Differences between the Swedish-Turkish bilinguals and the other groups were thus smaller when narrative vocabulary was assessed compared with vocabulary production on a test. Despite their low performance on the vocabulary task, the Swedish-Turkish bilinguals performed comparably in terms of NDW to the Swedish-German bilinguals in both MAIN-tasks and to the monolinguals in MAIN2. The results show that how vocabulary is assessed influences how groups perform in relation to each other. Using a type of measure such as NDW does thus not fully show the difference between the groups that is there when vocabulary knowledge is explicitly measured with a test. Explicit knowledge of words may not determine how well a child can tell a narrative to the same extent as the child's ability to use his/her vocabulary in the context of a narrative task. The influence of NDW and vocabulary production scores on narrative macrostructure is investigated in Chapter 8. For now, it is enough to point out that the fact that the type of measure influences results needs to be taken into account in further studies of vocabulary in bilinguals.

Age effects also varied for NDW in the two narrative tasks. For the Swedish MAIN1, there was only a difference between the six-year-olds who had a higher NDW and the younger groups. For MAIN2, on the other hand, there were differences between all three age groups: older children produced a higher NDW than younger children. In German, there was only an effect of age on MAIN2, whereas there were no age effects in Turkish. These results show that age is a more important factor for performance in the majority language than in the minority language, not only for scores on a vocabulary task, but also for narrative vocabulary. This is in line with the findings of Uccelli & Páez (2007) for NDW in Spanish-English bilinguals. The fact that no differences could be found between age 4 and 5 in the MAIN1 is similar to the findings of Muñoz et al. (2003) for Frog-story narratives in English of Spanish-English bilinguals. The fact that differences were found between the six-year-olds and the younger children on both narrative tasks in Swedish, just as was the case for vocabulary production, again suggests that a large step in development is taken between age 5 and 6. There is a clearer separation between the six-year-olds and the younger children than between age 4 and 5.

The second research question on NDW concerns the relationship between NDW and scores on vocabulary production (CLT). Vermeer (2000) found that NDW was a relatively good measure for distinguishing between children with different levels of lexical proficiency in Dutch, as it correlated with scores on two vocabulary tasks. The results for Swedish in the current study indicate that the relationship between expressive lexical knowledge as measured by a vocabulary task and lexical diversity in narratives is not the

same for monolinguals and for different bilingual groups. Whereas there were strong correlations between vocabulary production and NDW for the Swedish-Turkish bilinguals, the correlation was relatively weak for the monolinguals and not significant for the Swedish-German bilinguals. For the minority languages German and Turkish, there were strong correlations between NDW and CLT scores. For all three languages, results were the same for both MAIN1 and MAIN2. This indicates that, even taking the effect of age into account, there are other factors influencing NDW than lexical knowledge as measured by scores on a vocabulary task. Such factors could be the ability to carry out the narrative task, which is linked to the ability to interpret the pictures and understand the purpose of the task, and general fluency and grammatical proficiency in the language. It is unclear, however, why these aspects would lead to differences between languages and groups.

To conclude, the results from the analyses carried out for vocabulary point to important differences between the two bilingual groups, but these differences are especially pronounced in vocabulary production and much less so in narrative vocabulary. In vocabulary production, the Swedish-Turkish bilinguals lag behind their Swedish-German and monolingual peers considerably even at age 6, whereas they perform more similarly to the other groups in narrative vocabulary. Additionally, the relationship between vocabulary production (CLT) and narrative vocabulary (NDW) was not the same in the three language groups. It may be the case that a more thorough analysis of the use of different types of words would reveal further differences between the groups, differences that cannot be found when the broader measure of NDW is used. The differences in Swedish vocabulary between the Swedish-Turkish bilinguals on one hand and the monolinguals and Swedish-German bilinguals on the other are much larger when vocabulary production is measured using the test CLT than when NDW in the narratives is used as a proxy for lexical knowledge. Taken together, the results indicate that these different measures of vocabulary may not tap into the same type of lexical knowledge, and importantly, yield different results with regards to effects of age and language group.

6 Character introduction

Introducing referents in narrative discourse in a way that is understandable to a listener, who does not know the story or the character(s), is important for telling a good story; properly establishing referents is central for the continuation of the discourse. The ability to introduce referents appropriately is thus a central part of narrative ability. In this chapter, introductions of story characters in the MAIN1 (Cat/Dog) and ENNI narratives are analyzed.[82] Character introduction is defined as the first mention of a character in the narrative. The reason for performing the analysis on story characters and not on all referents is that story characters are the most essential referents, as they are agents of narrative events. It can therefore be assumed that they are included by most children in their narratives. Other referents, e.g. inanimate objects, are less prominent, especially in the stories studied here.

This part of the study is carried out on the data from the Swedish MAIN1 (Cat/Dog) and ENNI narratives for all three language groups (monolinguals, Swedish-German bilinguals, Swedish-Turkish bilinguals. Moreover, the Swedish-German bilinguals' performance in Swedish and German is compared. MAIN1 (Cat/Dog) and ENNI (A2/B2) were chosen for the analysis as they contain the same number of story characters. Swedish and German have comparable systems for introducing referents (see Section 6.2), using morphological (in)definiteness markers, whereas Turkish has a different type of system. Turkish has no definite article, and optional (and infrequently used) marking with an indefinite article. Instead, word order and case marking are used for marking the information status of a referent (cf. Küntay, 2002). Because the analyses performed concern the morphological form of the referring expressions used to introduce the story characters, results from Turkish are not reported here.[83]

To my knowledge, no study comparing the performance on referent introduction of different bilingual groups speaking the same majority language has been carried out. Therefore, the current study investigates character introduction in the Swedish narratives of Swedish-German and Swedish-

[82] For a study on anaphoric reference (i.e. referent reintroduction and maintenance), of the monolingual data from the current study, see Lindgren & Vogels (submitted). Results from the Swedish monolinguals' character introductions in MAIN1 and ENNI and from MAIN1 in both languages of 40 of the 46 Swedish-German bilinguals are reported in Lindgren (2018) and Reichardt (2016), respectively.

[83] An additional reason was that no data from ENNI had been collected in Turkish.

Turkish bilinguals. As German has a similar system as Swedish, using morphological markers of (in)definiteness, but Turkish a different one, using case markers and word order (see Section 6.2), comparing the performance of these two groups may allow us to draw conclusions about the role of cross-linguistic influence in bilinguals' production of referring expressions. Additionally, studying the performance in both languages of the Swedish-German bilinguals may lead to further insights into whether performance is different in two languages with comparable (in)definiteness-marking systems when one is the majority (societal) and one is a minority (home) language.

The aim of this chapter is to investigate which types of referring expressions are used for introducing story characters in the Swedish MAIN1 (Cat/Dog) and ENNI narratives and, for the same narratives, in both languages of the Swedish-German bilinguals. The following four specific questions are asked:

- Is there a development with age in the use of appropriate referring expressions for introducing story characters?
- For Swedish, are there any differences between the language groups?
- Do the Swedish-German bilinguals perform differently in their two languages?
- Is there any effect of the stimulus material, i.e. is there a difference in performance between MAIN1 (Cat/Dog) and ENNI (A2/B2)?

This chapter proceeds as follows. After summarizing findings from studies of children's referent introductions (Section 6.1), with subsections on monolinguals (Section 6.1.1)[84] and bilinguals (Section 6.1.2), and an overview of the Swedish, German and Turkish referential systems (Section 6.2), coding and analysis of the current study are described (Section 6.3). In Section 6.4, results are reported, first for Swedish (Section 6.4.1) and then for both languages of the Swedish-German bilinguals (Section 6.4.2). The chapter concludes with a summary of the results and discussion (Section 6.5).

6.1 Referent introduction in children

Being able to appropriately introduce referents in discourse requires knowledge of the linguistic means used for marking the information status of referents in the language, e.g. markers of novelty of information, be they morphological cues such as indefinite markers, or syntactic (clause/sentence level) ones such as word order (Chafe, 1976; Halliday & Hasan, 1976). Children need to understand the function of these markers, i.e. that certain

[84] This section is an extended version of the literature review in Lindgren (2018).

forms, e.g. indefinite noun phrases (NPs), are used to refer to objects and persons that are new to the listener, whereas other forms, e.g. definite noun phrases (NPs) and pronouns, are used for referents that have already been mentioned (Gundel & Johnson, 2013). Additionally, in order to use different types of expressions appropriately in specific discourse contexts, children must be able to correctly assess the listener's knowledge state, i.e. judge whether or not the referent is currently known to the listener (Ariel, 1990; Arnold, 2008; Gundel, 2010; Gundel, Hedberg, & Zacharski, 1993). To do this, they need to be able to take the listener's perspective into account, and be aware of the fact that it may differ from their own perspective. This means having developed sufficient Theory of Mind (ToM; Leslie, 1987; Leslie et al., 2004; Tomasello, 2003). Perspective-taking may also require additional executive function abilities, e.g. to be able to inhibit one's own perspective. (For an overview of the link between executive functions, ToM and referential abilities, see the review in De Cat, 2015). There are also indications that adequate reference in narrative discourse requires a well-developed capacity to update working memory as well as short-term memory (Whitely & Colozzo, 2013). Introducing referents appropriately is thus not a trivial task, especially not for bilinguals who need to develop the ability to introduce referents appropriately in both their languages, languages that may or may not have the same type of referential system(s).

6.1.1 Referent introduction in monolingual children

A relatively large number of studies have been carried out on the development of the ability to correctly introduce referents in monolingual children. Although the majority of the studies deal with English, a number of other languages have been investigated. However, little is known about Swedish-speaking children's referent introductions and when they learn to introduce referents appropriately in narrative discourse. With the exception of an unpublished MA thesis comparing Swedish monolingual and Swedish-English bilingual 6–7-year-olds by Finnstedt (2013) (see Section 6.1.2), the present study is the first to systematically investigate referent introductions for a larger group of Swedish-speaking children. The monolingual results have previously been reported in Lindgren (2018) and part of the Swedish-German data has been analyzed in the MA thesis by Reichardt (2016).

Mixed findings regarding the age at which monolingual children are able to introduce referents appropriately have been reported in the literature. Studies of spontaneous interaction indicate that already at age 2–3 children are able to use the full range of referring expressions correctly (Gundel & Johnson, 2013) and are also sensitive to discourse cues such as prior mention and joint attention with the interlocutor when choosing referring expressions (e.g. Hughes & Allen, 2013; see also the overview in Allen, Skarabela, & Hughes, 2008). For example, young children are more likely to omit new

referents when there is joint attention with the listener but use lexical forms when this is not the case (Skarabela, Allen, & Scott-Phillips, 2013). Results from a number of experimental studies also indicate that when producing referring expressions children take listener knowledge into account at an early age (Matthews, Lieven, Theakston, & Tomasello, 2006; Wittek & Tomasello, 2005). The tasks of those studies were relatively simple, involving for example a setup in which the children's production consisted of answers to specific questions from adults such as "What did you see?" (Matthews et al., 2006). Such tasks cannot be compared to more complex narrative tasks.

Results from studies of elicited narratives are inconclusive as to the exact age when children are able to introduce referents appropriately. Some studies report predominant use of appropriate indefinite expressions and low proportions of so-called egocentric definiteness errors, i.e. the use of definite expressions (pronouns and definite NPs) to introduce a referent that is unknown to the listener, already at age 2–4 (De Cat, 2013; Emslie & Stevenson, 1981). For example, Emslie & Stevenson (1981) found that English-speaking two-year-olds produced 72% indefinite NPs when introducing characters in two three-page stories. The three- and four-year-olds in their study produced more than 80% indefinites. Other studies indicate that the ability to introduce characters appropriately develops much later and may not be fully developed until age 9 (e.g. Hickmann, Hendriks, Roland, & Liang, 1996; Kail & Hickmann, 1992; Serratrice, 2007; Warden, 1976), or is at least not acquired before age 7 (Küntay, 2002; Schneider & Hayward, 2010).

In a well-known study of narratives elicited with the Frog story (Mayer, 1969), Hickmann et al. (1996) studied the introduction of animate characters in English, French, German and Mandarin Chinese narratives by monolingual preschool children (aged 4–5), seven- and ten-year-olds, and adults. The preschool children in this study did not use indefinite markers systematically to introduce characters. The authors therefore concluded that "despite some early uses of indefinite/definite determiners, the discourse-internal function of this opposition for the contrastive marking of new/given information is learned late" (Hickmann et al., 1996, p. 613). Similar results were reported in Schneider & Hayward (2010), who analyzed referent introductions using a First Mentions score based on referential adequacy in the narratives of 300 English-speaking children with typical language development and 77 children with language impairment aged 4–9. In this study, narratives were elicited with ENNI (Schneider et al., 2005), part of which is also used in the present study. In addition to differences between children with typical development and with language impairment, there were clear effects of age. The authors concluded that the children's ability to introduce referents "gradually improves until age 7, when it appears to be mastered by the majority of children" (Schneider & Hayward, 2010, p. 467).

One of the factors that may explain differences in the age at which referring expressions are used appropriately in narrative discourse is *language structure*. Some studies have found differences between languages in children's performance (e.g. Hickmann et al., 1996 for Mandarin Chinese, English, French and German). Investigating 3–5-year-old children's character introductions in Greek, English and Turkish, Aksu-Koç & Nicolopoulou (2015) point to the influence of the language's referential system on children's performance, and suggested that children speaking languages that use a rich morphological system to mark information status of referents (e.g. Greek) acquire appropriate use of the different forms earlier. The Greek-speaking children produced a higher proportion of appropriate referent introductions than the English-speaking children, who in their turn performed better than the Turkish-speaking children. The Greek children were the only group to produce a substantial proportion of indefinite NPs already at age 3, and this type was dominant by age 5, whereas the other children used indefinite NPs with a much lower frequency, or, in the case of the Turkish-speaking children, almost none at all. It should be noted that the Greek children produced a high proportion of indefinite NPs (around 50% for the five-year-olds) despite the fact that there was shared visual attention with the adult listener. This seems to indicate that in some languages, the ability to introduce referents appropriately develops earlier. However, results from other studies point to similar ages of acquisition across a range of different languages, for example French (Kail & Hickmann, 1992), Spanish (Kail & Sanchez y Lopez, 1997), Turkish (Küntay, 2002), Cantonese (Wong & Johnston, 2004), and Japanese (Nakamura, 1993).

Another factor influencing children's use of referring expressions is the type of *stimulus material* and the *elicitation method*. De Cat (2013) used materials in which "[t]he plots were very simple, and consisted essentially in the progressive introduction of new characters joining a group of established characters" (De Cat, 2013, p. 60), which may explain why the French-speaking children in her study used 83% indefinite NPs in introductions of story characters already at age 2;6–3;3. In contrast, studies finding that children do not introduce referents appropriately until at a later age generally used more complex story elicitation materials, such as the Frog story (Mayer, 1969), which is a long wordless picture book with a plot including several complications before the goal is achieved. Plot complexity (including the number of episodes), number and types of characters (e.g. animals or humans, main or auxiliary characters), and how easily characters are distinguishable from each other are aspects that also affect the child's ability to appropriately introduce referents (e.g. Aksu-Koç & Nicolopoulou, 2015; Colozzo & Whitely, 2014; De Cat, 2013; Hickmann et al., 1996; Wigglesworth, 1990). In the study by Aksu-Koç & Nicolopoulou (2015) described above, the children produced more indefinite noun phrases when they told a story with two main characters, compared to a story with one main and one

secondary character. In a study of Spanish-speaking children, Kail & Sanchez y Lopez (1997) found a strong effect of type of referent; secondary characters were predominantly introduced with indefinites by all age groups, but main characters were not.

To summarize the results on referent introductions discussed above, studies in which very young monolinguals performed well used elicitation instruments and procedures that were relatively simple, e.g. narrative stimuli with a simple episodic structure (or plot) and/or with few characters who are easily distinguishable from each other. In contrast, studies that found a late mastery of correct forms for referent introductions often used more complex story elicitation materials. Furthermore, some of the studies that report low production of indefinite NPs by relatively old children, e.g. seven-year-olds, used procedures with shared visual attention between child and experimenter.[85] It remains unclear which effect language structure has on the development of the ability to introduce referents appropriately in narrative discourse.

Relatively few studies have compared children's performance concerning referent introduction on different elicitation materials. Some studies have compared different elicitation procedures, e.g. with or without a blindfolded experimenter (Kail & Hickmann, 1992), used stories with one or more main characters (Aksu-Koç & Nicolopoulou, 2015; Colozzo & Whitely, 2014) or compared problem-based with event-based stories (Shapiro & Hudson, 1991). However, elicitation instruments were often similar in layout, length and/or story complexity. To fully investigate the effect of the stimulus, it is necessary to compare children's performance using narrative elicitation instruments that are more dissimilar, which is done in the current study.

Although studies vary in their results, there is ample indication that the late preschool years, i.e. age 4–6, are central for the development of the ability to introduce referents appropriately (Aksu-Koç & Nicolopoulou, 2015; Schneider & Hayward, 2010), just as is the case for the development of narrative ability more generally (Berman & Slobin, 1994c). The age range of the current study is therefore appropriate for being able to pinpoint when large steps in development take place (cf. Aksu-Koç & Nicolopoulou, 2015).

6.1.2 Referent introduction in bilingual children

Far fewer studies have been carried out on bilingual children's referent introduction than on monolinguals and results are not fully conclusive as to when the ability to introduce characters appropriately develops in bilinguals, and whether and how bilinguals differ from monolinguals. Moreover, reference in Swedish-speaking bilinguals has rarely been investigated. To my

[85] This has also been pointed out by De Cat (2013, p. 59) who writes: "the way in which the common ground is established has a significant impact on the child's ability to keep track of their interlocutor's perspective".

knowledge, the only study of reference in another group of Swedish-speaking bilinguals than those of the current study is the MA thesis by Finnstedt (2013). Finnstedt (2013) analyzed referent introduction and maintenance in 6–7-year-old Swedish monolinguals and in both languages of Swedish-English bilinguals.

Results from some earlier studies suggest that bilingual children perform similarly to monolingual children in referent introduction. For example, in a study of referent introduction, re-introduction and maintenance in Frog story narratives, Serratrice (2007) found that Italian-English children aged 6;11–9;11 performed similarly to monolinguals in both languages. Similarly, the study by Finnstedt (2013) showed no difference between Swedish monolinguals and Swedish-English bilinguals with 75–78% appropriate referent introductions in both groups of 6–7-year-olds. The Swedish-English bilinguals performed similarly in their two languages.

There are also studies that show that bilinguals develop slightly differently from monolinguals in one or both languages with respect to referent introduction, or that they do not have comparable performance in both languages (Álvarez, 2003; Chen & Lei, 2013; Chen & Pan, 2009; Jia & Paradis, 2015; Reichardt, 2014). For example, Chen & Lei (2013) found that 9-year-old Chinese-English bilinguals living in the U.S. produced lower proportions of indefinite NPs to introduce characters in their English narratives compared with English monolinguals, but performed comparably to monolinguals in Chinese. Jia & Paradis (2015) found that, when introducing characters and objects in narratives in Chinese, Chinese-English bilinguals aged 6;9-10;10 growing up in Canada used fewer appropriate referring expressions than Chinese monolinguals.

In a study of Russian-German bilinguals aged 4;0-6;11, Topaj (2010) found that the bilinguals were comparable to monolinguals in both languages if all referring expressions that were sentence topics were considered (the bilinguals were compared to the Russian and German monolinguals from Gülzow & Gagarina, 2007); types of referring expressions for introducing characters specifically were not compared. However, the bilinguals of Topaj (2010) produced only 35% indefinite NPs for new referents (i.e. referent introduction) in German. This can be compared with 50% and 65% indefinites for 4–5-year-olds and 7-year-olds, respectively, of the German monolinguals in the study by Hickmann et al. (1996). In a study of German-Russian 4-, 7- and 9-year-olds, Reichardt (2014) found that 9-year-olds performed similarly to the monolingual 7-year-olds from Hickmann et al. (1996) and that 4- and 7-year-olds rarely used indefinite NPs to introduce characters. The results indicate that German-Russian bilinguals develop the ability to use indefinite NPs appropriately later than German monolinguals.

In sum, results from earlier studies indicate that bilinguals may not perform similarly to monolinguals in both languages when the languages have different types of referential systems, such as English and Chinese (Chen &

Lei, 2013; Chen & Pan, 2009; Jia & Paradis, 2015) or Russian and German (Reichardt, 2014; Topaj, 2010), but that bilinguals speaking languages with similar/comparable systems, e.g. Swedish and English (Finnstedt, 2013) or English and Italian (Serratrice, 2007), may perform similarly to monolinguals.

Different groups of bilinguals with the same combination of languages may also perform differently. In their study of 8–12-year-old German-Greek bilinguals, Andreou, Knopp, Bongartz & Tsimpli (2015) found differences between bilinguals and monolinguals in Greek, but not in German. In German, both bilinguals living in Germany and those living in Greece performed similarly to monolinguals. In Greek, bilinguals living in Germany produced substantially fewer indefinite NPs than bilinguals living in Greece and monolinguals. As the referential systems of Greek and German are similar, the authors interpret this as a "delay imposed by exposure effects" (Andreou et al., p. 43), i.e. having less exposure to a language may mean that the process of learning how to introduce characters appropriately is slower. It remains unclear why Greek was affected to a larger extent than German. Jia & Paradis (2015) found that Chinese-English bilinguals who arrived in Canada at a later age and/or had mothers with higher education performed better on referent introduction in Chinese. For children who attended English-only schools, but not for children who attended bilingual schools, the richness of Chinese input was also a significant factor. The results from earlier studies indicate that bilinguals' performance may be influenced by a number of different factors in addition to age, such as the language combination, language proficiency and the amount of exposure.

6.2 The Swedish, German and Turkish referential systems

The systems for marking the information status of a referent in Swedish and German are similar. Both languages mark (in)definiteness and specificity morphologically, and so-called global markers, such as word order, do not play a role. Turkish, on the other hand has a different type of system, employing case markers and word order to indicate the status of a referent. In contrast to Swedish and German (as well as English), Turkish does not mark definiteness morphologically at all, and the indefiniteness marker is optional (see e.g. Küntay, 2002). Additionally, whereas Swedish and German have different third person singular pronouns, Turkish has just one. Turkish information status marking is thus substantially different from Swedish and German. An overview of Swedish, German and Turkish forms of reference, with a comparison to English is given in Table 6.1.

Table 6.1. The Swedish, German and Turkish referential systems.

	Indefinite NP	Definite NP	Pronoun
Swedish	en pojke	pojken	han / hon / den/det / dom
German	ein Junge	der Junge	er/der / sie/die / es/das / sie
Turkish	(bir) çocuk	çocuk	o / onlar
English	a boy	the boy	he / she / it / they

Note. Nominative case only.

In Swedish, the indefiniteness marker is a freestanding article marked for gender (*en* and *ett* for common and neuter gender respectively). The Swedish definiteness marker is a suffix, marked for gender (*-en*/*-et*). Although the definiteness suffixes are acquired earlier in Swedish than the indefinite articles, i.e. in spontaneous speech Swedish-speaking children mark definiteness earlier than indefiniteness (Bohnacker, 2003, 2007; Kupisch, Anderssen, Bohnacker, & Snape, 2009), both types of morphological markers emerge early in children's spontaneous speech, around or even before age 2.[86]

Swedish has four singular pronouns, two marking the biological gender of animate referents (*hon* 'she' and *han* 'he') and two marking the two grammatical genders common and neuter (*den*/*det* 'it'), and one plural pronoun (*dom* 'they).[87] The two pronouns used for the two grammatical genders can be used for animate referents as well (most animate nouns belong to the common gender). Swedish pronouns have object forms, making them the only referring expressions in the language that are inflected for a case other than the genitive.

The difference between the Swedish and the German indefinite articles and pronouns is that the German ones are inflected not only for gender, but also for case. German, like English, has a definite article, which is marked for gender and case in a similar way to the indefinite article. German (in)definiteness markers have been found to be acquired at a similar age to the Swedish ones (Gülzow, in preparation; Kupisch et al., 2009).

The discourse-pragmatic rules for the use of the different forms are the same for Swedish and German (cf. e.g. Helbig & Buscha, 2001; Teleman, Hellberg, & Andersson, 1999: 269ff).[88] When there is no shared knowledge or joint visual attention between speaker and listener, an indefinite NP, as in (1) and (2), is normally used to introduce a new referent appropriately in

[86] That morphological markers are present in the child's speech does not mean that they are used correctly for different information statuses in different contexts and/or types of discourse.

[87] In written Swedish, this pronoun appears as *de* for subject and *dem* for object case, but both these forms are pronounced as *dom* in most varieties. Therefore, the form *dom* is used in all examples.

[88] Note that the semantic rules for the use of bare nouns are slightly different. Bare nouns are more widely used in Swedish than in German (cf. Bohnacker, 2003, 2007).

both Swedish and German (see Helbig & Buscha, 2001; Teleman, Hellberg & Andersson 1999: 169ff).[89]

(1) det var en elefant och en giraff som skulle bada
'There were an elephant and a giraffe that were going to swim.'
(MoSwe6-11, 6;4, A2)
(2) es war einmal eine Katze
'Once upon a time, there was a cat.' (BiGer6-08, 6;11 – Cat)

Using a definite form to introduce a character, as in (3) and (4), gives the listener enough lexical information to understand who the story is about. However, it signals that the referent is uniquely identifiable in the context (Gundel et al., 1993), i.e. that it can be seen by speaker and listener or is in another way already known to the listener.

(3) katten tar en fjäril
'The cat catches a butterfly.' (MoSwe4-02, 4;5 – Cat)
(4) dann fängt der Junge der Ballon
'Then the boy catches the balloon.' (BiGer5-06, 5;2 – Dog)

If a pronoun is used to introduce a character, as in (5) and (6), it is impossible for the listener to know who the character is. Pronouns are clearly less felicitous than any type of lexical NP, as using a pronoun not only indicates that the referent has been mentioned earlier in the discourse, but also that it is activated in the listener's mind (cf. Gundel et al., 1993).

(5) att han ska ta en fjäril
'that he is going to catch a butterfly.' (MoSwe5-04, 5;10 – Cat)
(6) die jagt den Schmetterling
'She/it chases the butterfly.' (BiGer4-10, 4;1 – Cat)

Swedish and German thus employ morphological forms (cf. Table 6.1) to mark the information status of the referent (cf. e.g. the Givenness Hierarchy, Gundel et al., 1993; see also Ariel, 1990; Givón, 1983). As data from Turkish are not analyzed here, the Turkish system is not described in detail (see Küntay, 2002 for an overview), but in Turkish a lexical NP would also be used to introduce a referent; pronouns or null forms would not be acceptable for referent introduction. To summarize, Table 6.2 gives an overview of the preferred forms for referent introduction in Swedish, German and Turkish.

Table 6.2. Preferred form for referent introduction in Swedish, German and Turkish.

	Swedish	**German**	**Turkish**
Preferred form for referent introduction	en pojke 'a boy'	ein Junge 'a boy'	çocuk 'boy'

[89] In all examples in this chapter, hesitations, repetitions and false starts have been removed for increased readability. The introduced character is underlined.

Being the least appropriate and most appropriate types of referring expressions for introducing new referents, pronouns and indefinite NPs form the extreme ends of the (accessibility) scale in both Swedish and German (as well as in other languages that have indefinite NPs and do not employ null forms). Therefore, pronouns and indefinite NPs are the most interesting types of referring expressions to analyze, and will be the focus of the analyses carried out in this chapter.

6.3 Coding and analysis

The analyses for Swedish and German reported in this chapter were carried out on the MAIN1 (Cat/Dog) and ENNI (A2/B2) data. The reason for choosing these two narrative tasks for the analysis is that they are comparable in terms of number of characters; each of the four stories (Cat, Dog, A2 and B2) has three characters.[90] Table 6.3 shows the story characters in MAIN1 and ENNI, respectively.[91]

Table 6.3. Overview of the MAIN1 and ENNI story characters.

Narrative task	**MAIN1**		**ENNI**	
Story	**Cat**	**Dog**	**A2**	**B2**
Character 1	Cat	Dog	Small elephant	Small rabbit
Character 2	Butterfly	Mouse	Giraffe	Dog
Character 3	Boy	Boy	Large elephant (Lifeguard)	Large rabbit (Doctor)

Two Swedish narratives from each child and two narratives in German from each Swedish-German bilingual were analyzed, with a few exceptions due to missing data. There were no Swedish MAIN1 data from one Swedish-German five-year-old, and no German ENNI data from two Swedish-German children, as well as no Swedish ENNI data from one Swedish-German bilingual four-year-old and four Swedish-Turkish bilinguals. Additionally, ENNI-data from one Swedish-Turkish six-year-old (BiTur6-12) had to be discarded because the experimenter diverted from task procedure by looking at the pictures together with the child. It was not meaningful to analyze character introduction in the resulting narrative, as it would not be com-

[90] The MAIN2 stories (Baby Birds/Baby Goats) have five characters and are thus not comparable to MAIN1 and ENNI.

[91] Table 6.3 serves as a reminder of the characters depicted in the pictures (see Section 3.2). The possible effect of type of character on referent introductions is not analyzed here. For a study of the effect of animacy on referent reintroduction and maintenance in the Swedish MAIN1 monolingual data, see Lindgren & Vogels (submitted), and on character introduction in MAIN1 from 40 of the 46 Swedish-German children, see Reichardt (2016).

parable to the other narratives.[92] In sum, Swedish data from 165 children for MAIN1 and from 160 children for ENNI were analyzed (cf. Chapter 4). German data were analyzed from 46 children for MAIN1 and 44 children for ENNI. A total of 325 narratives in Swedish and 90 narratives in German were thus analyzed.

6.3.1 Coding and categorization of referring expressions

Each referring expression used to introduce a story character, i.e. the first expression that was used to refer to that character, was identified in the narratives. The maximum number of expressions per child was three per narrative, one for each story character. Each expression was coded according to type of referring expression (e.g. pronoun, indefinite or definite NP), language group (Swedish monolingual, Swedish-German bilingual, Swedish-Turkish bilingual; for the analysis of Swedish only), age group (four-, five- or six-year-old), narrative task (MAIN, ENNI), story (Cat, Dog, A2, B2), story character (see Table 6.3), and language (German, Swedish; only for the data from the Swedish-German bilinguals). To clarify, coding according to type of referring expression means that expressions were coded according to their lexical and morphological form. Whether or not the child used correct gender marking, and in the case of German also case marking, was not taken into account when coding the expressions according to type. This was done so as to not punish the bilingual children for mistakes related to the gender of the noun; the aim was not to judge the children's knowledge of the gender systems, but to assess their ability to use referring expressions appropriately, i.e. to use markers of (in)definiteness for introducing characters.[93] As long as the child used an overt (morphological) (in)definiteness marker (in the target language) the expression was coded as an indefinite or definite NP.

Personal pronouns (both singular and plural), demonstrative pronouns, e.g. *den där* 'that', and indefinite pronouns, e.g. *en* 'one' and *nån* 'someone' (see (15) and (16) below) were included in the pronoun category. Also coded

[92] There were also some cases in which the experimenter used prompts that were not allowed according to protocol: In the ENNI testings of five Swedish-Turkish bilinguals (one four-year-old, one five-year-old, and three six-year-olds), the child started the narrative by introducing one or two of the characters using a pronoun, and the experimenter then asked *vem är det?* 'who is it?'. The child's first reference to the character(s), i.e. the pronoun, was counted as an introduction, but it is possible that the fact that the experimenter asked this question may have influenced the child's subsequent character introductions, making them more appropriate than would otherwise have been the case.

[93] In the Swedish data from the Swedish-Turkish bilinguals, there was a relatively high proportion of gender-related mistakes. Mistakes with gender were also fairly common in the German data of the Swedish-German bilinguals. In the German narratives, the most notable was the overuse of the article *den* 'the.MASC.ACC' which was used frequently by some of the Swedish-German children for other genders and/or cases, possibly due to transfer from Swedish, where the homophonous pronoun and definite article *den* 'it/the' used for the common gender is generally very frequent.

as pronouns were a few possessive pronouns (those relating to a character not previously mentioned in the child's narrative), e.g. in utterances such as (7), where the overgeneralized possessive pronoun *hons* 'she's' (the correct form is *hennes* 'her') refers to the boy, who has not been mentioned previously in the child's narrative.

(7) sen hons ballong är i vattnet
'Then she's balloon is in the water.' (BiTur4-07, 4;4 – Cat)

Coding the character introductions was in most cases straightforward. Even for pronouns it was always possible to identify which character they referred to by looking at the content of the accompanying clause and comparing it with the actions performed by the different characters in the stories (i.e. in the pictures).[94] Only a few of the referring expressions produced in the Swedish narratives require further comment with regard to how they were coded and why.

One character was introduced with an ungrammatical null form, shown in (8). This expression was produced by a Swedish-Turkish four-year-old to introduce the boy.[95] Both null forms and pronouns are used for highly salient, given referents in adult speech (cf. Ariel, 1990) and it thus makes sense to put this single case of a null form together with the pronouns for the analyses.

(8) sen ta den där, ballongen
'Then (he) take(s) that one, the balloon.' (BiTur4-29, 4;2 – Dog)

One Swedish-Turkish five-year-old introduced the mouse in MAIN1 Dog with an indefinite article combined with a demonstrative pronoun, forming the non-idiomatic expression *en dom där* 'one those there'. The child may have intended, but not fully mastered the similar sounding expression *en sån där* 'one like that/one such thing'. Although this referring expression contains the indefinite article *en* 'a', it does not give any further lexical information, i.e. there is no noun, and it is therefore not more informative than a regular pronoun. It was therefore coded as pronoun.

A similar yet different case was found in the Swedish data of the Swedish-German children. The child introduced the character (the mouse) with a

[94] A few referring expressions were produced in isolation, i.e. as free-standing NPs without a clause. In all cases, these were lexical NPs (indefinite, definite or bare nouns), and thus the referent introduced by these expressions could be identified.

[95] As the verb form produced is the infinitive *ta* 'to take', it is not completely clear if the utterance should be interpreted as present tense with a left-out present tense ending (present tense *tar* 'takes'). It could also be the case that an auxiliary verb, e.g. *ska* 'is going to' has been left out, and that the utterance is supposed to mean that the boy is going to take the balloon. Thus, the translation given here is only one possible interpretation. However, it is clear that the utterance describes an action of the boy and that it thus should count as introduction of the boy.

referring expression consisting of an indefinite article and her own idiosyncratic version of the noun *mus* 'mouse', a doubling of the stem with an added definite suffix, resulting in the (non-existing) form *en musmusen* 'a mouse-mouse-DEF'. The question here was whether to code this expression as indefinite or definite, as it contains both markers. After careful consideration it was decided to code this as an indefinite NP, as the made-up word *musmusen* may be seen as an unanalyzed chunk. To code this as definite would mean ignoring the child's correct use of the indefinite article.

Nine referring expressions produced in the German narratives (four cases in MAIN1; five cases in ENNI) were removed from the data and are thus not included in the analyses reported below. These cases consisted of referring expressions that were either completely in Swedish (four expressions) or partly so (five expressions), e.g. *hunden* 'the dog' (BiGer4-11, 4;5, Dog), *doktorn* 'the doctor' (BiGer4-18, 4;5, B2), *eine pojke* 'a [ger.] boy [swe.]' (BiGer4-15, 4;9, Dog), and *der kompis* 'the [ger.] friend [swe.]' (BiGer5-12, 5;8, B2).

The following types of referring expressions were found in the data: pronouns, definite NPs, indefinite NPs, bare nouns (i.e. nouns without any inflection/(in)definiteness marker, e.g. *katt* 'cat'), proper nouns (e.g. *Pappa Elefant* 'Daddy Elephant'), and possessive NPs. Possessive NPs were those possessive constructions that related the introduced character to a previously mentioned character. An example of a possessive NP is the introduction of the large rabbit as *kaninens mamma* 'the rabbit's mother' when the small rabbit had already been introduced.[96]

Finally, when coding of the morphological form of the referring expressions used to introduce story characters was completed, each referring expression was classified as either fully appropriate or not. In addition to indefinite NPs, proper nouns and possessive NPs were included in fully appropriate NPs. However, the category fully appropriate NPs mainly consisted of indefinite NPs, as proper nouns and possessive NPs were not common in the data (and in fact almost completely absent from the MAIN1 data). All other referring expressions were classified as not fully appropriate NPs.

6.3.2 Statistical analysis

The statistical analysis serves to answer the research questions about age development, and differences between stimulus materials, language groups, and, for the Swedish-German bilinguals, languages.

Chi-Square tests were carried out to ensure that there were no differences in the distribution of different types of referring expressions between the two

[96] This is very different from introducing the character in the form of a possessive *pronoun* as in (7) – in a possessive NP, the character is introduced with a noun, as 'the possessed', and not with a possessive pronoun as 'the possessor' (see also (9)).

stories of each narrative task, i.e. between Cat and Dog for MAIN1 and A2 and B2 for ENNI. These tests were run for each narrative task separately for the Swedish and the German data. Additionally, for the Swedish-German data, a Chi-Square test was run to test whether there were any differences between expressions produced in the first and in the second testing.

On both the Swedish data and the German and Swedish data of the Swedish-German bilinguals, two logistic regression analyses were carried out. The first analysis had pronoun, the least appropriate type of referring expressions, versus lexical NP (i.e. all referring expressions containing a noun) as the dependent variable. In the second analysis, the dependent variable was fully appropriate NP versus other NP (i.e. all other referring expressions).

In Swedish, the variables *age group*, *language group*, and *narrative task*, and the interactions between them, were predictors in both analyses. The variable narrative task was a simple binary predictor, comparing MAIN1 and ENNI. For the variable age group, one binary predictor compared the six-year-olds to the younger groups, and another one compared the five-year-olds to the four-year-olds (reversed Helmert coding).[97] The variable language group was coded into one binary predictor comparing the monolinguals (coded as 1) with all bilinguals (coded as 2 and 3), and another one comparing the Swedish-German bilinguals with the Swedish-Turkish bilinguals (Helmert coding). This coding of the language group variable is consistent with a focus on analyzing differences between monolinguals and bilinguals on the one hand and differences between the two bilingual groups on the other. The analyses of the data from the Swedish-German bilinguals had *age group*, *narrative task* and *language* (Swedish or German) and their interactions as predictors.

Model selection was used to find the simplest model with the best fit. This was done through (manual) stepwise removal of predictors with the highest p-values, starting with the higher-level interactions (i.e. first removing the three-way interaction and then the two-way interaction with the highest (non-significant) p-value and so forth). The models with and without the predictor were then compared to see if keeping the predictor significantly improved the model. The starting point was always the full model (i.e. a model containing main effects, all two-way interactions and the three-way interaction). For a description and justification of this method of model selection, see Field (2013, p. 768).

[97] As a logistic regression analysis requires all predictors to be binary, it is not possible to compare all the three groups to each other in one model (cf. Field, 2013, pp. 446–447 on contrasts).

6.4 Results

6.4.1 Swedish

In this section, results for character introductions in the Swedish narratives are presented for the MAIN1 (Cat/Dog) and ENNI (A2/B2) for the three language groups: the Swedish monolinguals, the Swedish-German bilinguals and the Swedish-Turkish bilinguals. The number of Swedish referring expressions used by the children to introduce story characters in the MAIN1 and the ENNI narratives is shown in Table 6.4.

Table 6.4. Total number of Swedish referring expressions used for introducing the story characters in MAIN1 (Cat/Dog) and ENNI (A2/B2), by language group.

	MAIN1	**ENNI**	**Total**
Monolinguals (N=72)	208	195	403
Swedish-German bilinguals (N=46)	131	118	249
Swedish-Turkish bilinguals (N=48)	136	110	246
Total	**475**	**423**	**898**

As shown in Table 6.4, 898 referring expressions used for introducing story characters were found in the Swedish MAIN1 (Cat/Dog) and ENNI (A2/B2) narratives of the mono- and bilingual children. For all groups, a somewhat lower number of referring expressions was found in the ENNI narratives. This was not because fewer characters were introduced by the children in the ENNI narratives, but because of a tendency for some children to introduce the two first ENNI-characters with a plural expression (e.g. *två kaniner* 'two rabbits', *dom* 'they' or *dom två* 'those/they two'). The mean number of characters introduced by each child (for all children) was 2.9 for both MAIN1 and ENNI (median for both stories: 3), indicating that most children introduced all characters. Table 6.5 shows the number of referring expressions per child for the three language groups.[98]

Table 6.5. Mean number of Swedish referring expressions per child in MAIN1 (Cat/Dog) and ENNI (A2/B2), by language group.

	MAIN1	**ENNI**
Monolinguals (N=72)	2.9	2.7
Swedish-German bilinguals (N=46)	2.9	2.6
Swedish-Turkish bilinguals (N=48)	2.8	2.6
Total	**2.9**	**2.6**

[98] Note that the number of referring expressions per child is not the same as the number of characters introduced, as more than one character can be introduced with one referring expression. The analyses below are carried out on referring expressions, and not on characters introduced.

Generally, the children in the younger groups produced somewhat fewer referring expressions; patterns for the three language groups were similar.[99]

6.4.1.1 Types of referring expressions

Table 6.6 gives an overview of the types of referring expressions found in MAIN1 for the three language groups. There were no significant differences between Cat and Dog in the distributions of types of referring expressions ($\chi^2(4, N = 475) = 3.069$, $p = .546$). Therefore, all analyses were carried out on the MAIN1 data as a whole.

Table 6.6. Types of referring expressions used to introduce story characters, percentages (%), Swedish MAIN1 (Cat/Dog), by language group.

	Swedish monolinguals (N=208)	**Swedish-German bilinguals (N=131)**	**Swedish-Turkish bilinguals (N=136)**
Indefinite NPs	65.4	74.0	50.0
Possessive NPs	0.5	0.0	0.0
Bare Nouns	4.8	4.6	5.9
Definite NPs	21.6	16.0	33.1
Pronouns	7.7	5.3	11.0
Total	**100.0**	**100.0**	**100.0**

Note. N = total number of referring expressions.

In all three language groups, proportions of pronouns were generally low, although somewhat higher in the Swedish-Turkish group than in the other two groups. Bare nouns were used in around 5% of the cases. In all language groups, the younger children produced a higher proportion of bare nouns (see Appendix 4, Table A4.3). In fact, only four cases of bare nouns were produced by six-year-olds (one case each by the monolingual and the Swedish-German six-year-olds; two cases by the Swedish-Turkish six-year-olds). Only one case of a possessive NP was found in the data from MAIN1. This consisted of the referring expression *hans gubbe* 'his man' (MoSwe4-16, 4;0) to introduce the boy in MAIN1 Dog. The possessive pronoun referred back to the already introduced dog. Although this (possessive) NP is not indefinite, it is as appropriate an introduction as an indefinite NP. When one character has already been introduced, it is fully appropriate to introduce a second character using an expression that shows a relation to the first character (cf. Fraurud, 1996).

The most frequent types of referring expressions across the board were definite and indefinite NPs, but there are some differences in the distribution of these categories between the language groups. The Swedish-German bilinguals produced the highest proportion of (fully appropriate) indefinite

[99] For a description of the data by age group, including the number of referring expression per child in each age group, see Appendix 4, Tables A4.1-A4.2.

NPs; almost three-fourths (74%) of their character introductions were indefinite. In comparison, the Swedish-Turkish bilinguals produced only 50% indefinite NPs, a full third less. The results from the monolinguals were in between. These differences in distributions are mirrored in the proportions of definite NPs.

In Table 6.7, the distributions of different types of referring expressions used for character introduction in the ENNI narratives are shown for the three language groups. No statistically significant differences were found between the A2 and the B2 on the distributions of different types of referring expressions ($\chi^2(5, N = 423) = 7.713$, $p = .173$). Data from the two ENNI-stories have therefore been collapsed in all analyses.

Table 6.7. Types of referring expressions used to introduce story characters, percentages (%) Swedish ENNI (A2/B2), by language group.

	Swedish Monolinguals (N=195)	**Swedish-German bilinguals (N=118)**	**Swedish-Turkish bilinguals (N=110)**
Indefinite NPs	39.0	36.4	21.8
Possessive NPs	5.6	3.4	5.5
Proper Nouns	3.1	1.7	0.9
Bare Nouns	4.6	4.2	11.8
Definite NPs	26.2	39.8	29.1
Pronouns	21.5	14.4	30.9
Total	**100.0**	**100.0**	**100.0**

Note. N = number of referring expressions.

The results for ENNI shown in Table 6.7 are strikingly different from the results presented for MAIN1 above (Table 6.6). In all language groups, proportions of pronouns were about three times as high in ENNI compared with MAIN1. Proportions of indefinite NPs are also around half the size in ENNI. Some of the difference in the proportions of indefinite NPs, but far from all, could be explained by the occurrence of possessive NPs and proper nouns in the ENNI data not found in the MAIN1 data. In all three groups, possessive NPs were more common than proper nouns, and the monolinguals produced a higher proportion of proper nouns than the other groups. Both of these categories were mostly used to introduce the third character, the lifeguard and the doctor in A2 and B2, respectively.[100] The main reason for this was that a number of children took the lifeguard, being a large elephant, and the doctor, being a large rabbit, to be the parent of the elephant/rabbit. This interpretation led the children to introduce them as e.g. *kaninens pappa* 'the rabbit's father', as in (9), or as *Mamma Kanin* 'Mother Rabbit', as in (10). The tendency to interpret relationships between the ENNI story characters

[100] In fact, 14 of the 21 possessive NPs and eight out of the nine proper nouns were used to introduce these two characters. They were evenly distributed between the lifeguard and the doctor.

in this way can also be seen in the relatively frequent use of the nouns *mamma* 'mother' and *pappa* 'father' to introduce the lifeguard and the doctor, as in (11). As these kinship nouns are often used in definite form, they may have contributed to the larger proportion of definite NPs in the ENNI compared with MAIN. However, they cannot explain the higher proportions of pronouns in the ENNI data.

(9) sen så kom kaninens pappa
'Then the rabbit's father came.' (MoSwe6-15, 6;5 – B2)
(10) och nu kommer Mamma Kanin
'And now Mother Rabbit comes.' (BiGer4-11, 4;5 – B2)
(11) och sen kommer mamman snabbt
'And then the mother quickly comes.' (BiGer4-05, 4;4 – A2)

In ENNI, the Swedish-Turkish bilinguals used a higher proportion of bare nouns than the other two groups, but this difference between the language groups is only found in the two younger groups (see Appendix 4, Table A4.4). The Swedish-Turkish six-year-olds used a low number of bare nouns, comparable to the six-year-olds of the other language groups. The Swedish-Turkish four- and five-year-olds, on the other hand, used much higher proportions of bare nouns than children their age in the other language groups. A more detailed analysis revealed a specific preference for using a bare form of the noun *doktor* 'doctor' for introducing the doctor in ENNI B2 among the Swedish-Turkish four- and five-year-olds.[101] In fact, eight of the 13 bare nouns were introductions of this character. The monolinguals and the Swedish-German bilinguals did not show this type of preference.

The monolinguals and the Swedish-German bilinguals produced similar proportions of indefinite NPs, whereas the proportion in the Swedish-Turkish group is substantially lower. In contrast to this, the monolinguals and the Swedish-Turkish bilinguals used similar proportions of definite NPs whereas the proportion in the Swedish-German group is higher. This is related to the fact that the Swedish-German group had a lower proportion of pronouns than the other two groups.

When comparing MAIN1 and ENNI, the results for the two narrative tasks are substantially different. An overview of types of referring expressions for MAIN1 and ENNI of the three language groups is given in Figure 6.1. Figure 6.1 clearly shows that, in all language groups, proportions of pronouns are higher and proportions of fully appropriate NPs are lower in ENNI compared with MAIN.

[101] Note that using a bare form corresponds to how the character would be introduced (and referred back to as well) in Turkish.

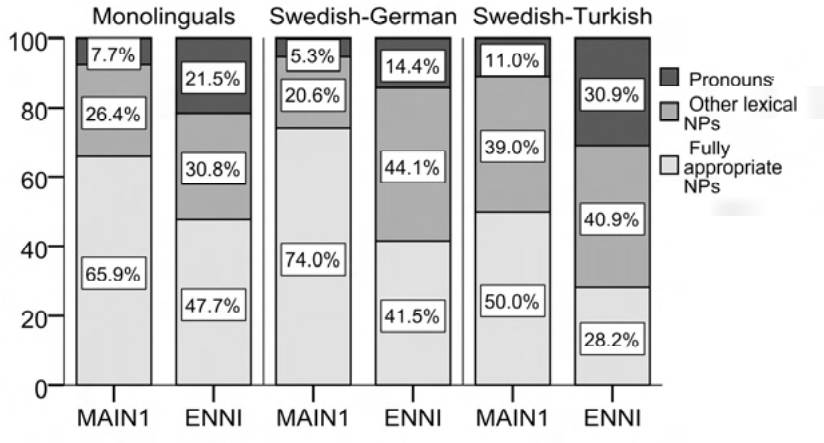

Figure 6.1. Character introductions, proportion (%) fully appropriate NPs, other lexical NPs, and pronouns by language group and narrative task, Swedish MAIN1 (Cat/Dog) and ENNI (A2/B2).

Differences between the age groups concerning other types of referring expressions than pronouns and fully appropriate NPs are not shown here. For the interested reader, detailed tables (including raw figures) for all different types of referring expressions used for character introductions by age and language group can be found in Appendix 4, Table A4.3-A4.4.

6.4.1.2 Pronouns

In Figure 6.2, the proportions of pronouns in Swedish MAIN1 are shown.

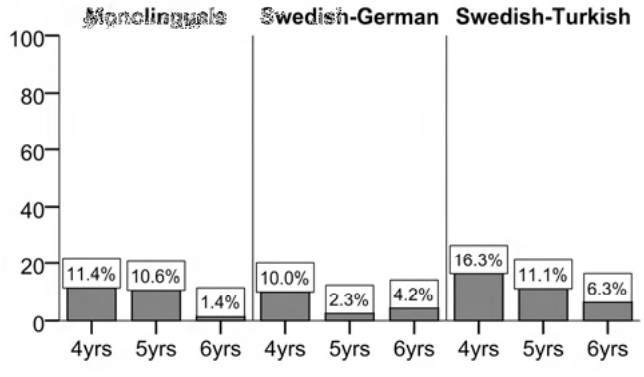

Figure 6.2. Character introductions, proportion (%) of pronouns by age and language group, Swedish MAIN1 (Cat/Dog).

As shown in Figure 6.2, only a small part of the referring expressions used to introduce characters in MAIN1 are pronouns. In all three language groups, the highest proportion of pronouns was produced by the youngest group, the four-year-olds, with a somewhat higher proportion produced by the Swe-

dish-Turkish group. Whereas the monolingual five-year-olds produced a similar proportion of pronouns as their four-year-old counterparts, the proportion of pronouns in both bilingual five-year-old groups is lower compared with their four-year-old peers. Typical examples of pronouns produced by four- and five-year-old children to introduce characters are given in (12), (13), and (14), all introducing the boy.[102]

(12) han hade fiskat små fiskar
'He had fished small fishes.' (MoSwe4-05, 4;8 – Cat)
(13) och han tar sin ballong
'And he takes his balloon.' (BiGer4-10, 4;1 – Dog)
(14) hon går hem hos sig
'She walks home to her place.' (BiTur4-01, 4;11 – Dog)

The six-year-olds used very few pronouns for character introduction. Only one pronoun was found in the monolingual six-year-old MAIN1 data. This case, shown in (15), is also somewhat special, as it is an indefinite pronoun (*en* 'one') with a relative clause that gives a description of the character, making it a more appropriate introduction than the pronoun introductions shown in (12), (13), and (14). A similar example in (16), involving another indefinite pronoun (*nån* 'someone') and adding information in the form of a modifier with *med* 'with' was produced by a Swedish-German six-year-old. This introduction was one out of only two pronouns produced by the Swedish-German six-year-olds. The Swedish-Turkish six-year-olds produced only three pronouns. These three cases were all similar to the pronouns produced by the younger children, as in (17).

(15) fast en som fiskade, han stoppade katten
'But one who was fishing, he stopped the cat.' (MoSwe6-01, 6;6 – Cat)
(16) kom nån med en ballong och korv i en kasse
'Someone came with a balloon and sausage in a bag.' (BiGer6-11, 6;2 – Dog)
(17) han var hungrig
'He was hungry.' (BiTur6-16, 6;4 – Dog)

To summarize, irrespective of language group, all children used relatively low proportions of pronouns to introduce the story characters in MAIN1; lexical NPs clearly dominated. Although the Swedish-Turkish group used somewhat higher proportions of pronouns, age effects in the three language

[102] Note that, in (14), the child BiTur4-01 refers to the boy as *hon* 'she'. This type of mistake, which is only found in monolingual children much younger than the ages studied here (and even then it is rare), is relatively common in the Swedish-Turkish data, especially in the younger age groups, and especially in children with relatively low Swedish proficiency. Additionally, this child's utterance as a whole is not fully idiomatic. Although the expression *hos sig* 'at one's own place' is idiomatic, it can only be used to denote a location. With a directional expression such as *gå hem* 'walk/go home', the reflexive *sig* 'oneself' needs to be combined with *till* 'to', forming the idiomatic expression *gå hem till sig* 'go home to one's own place'.

groups seem relatively similar, with lower proportions of pronouns in the older children. The number of pronouns used by the six-year-olds in MAIN1 was very low in all three language groups.

We now move on to results for pronouns in ENNI. First, some additional comments on pronouns in ENNI are warranted. Plural pronouns only occurred in ENNI. These pronouns, as in (18), were mostly used to introduce the first two characters (rabbit/dog or elephant/giraffe).

(18) dom vill bada
'They want to swim.' (MoSwe4-24, 4;2 – A2)

As in MAIN, some indefinite pronouns were produced by the older children. An example of a character introduction with an indefinite pronoun and relative clause construction is shown in (19).

(19) sen kommer en som jobbar i badet
'Then someone who works at the swimming pool comes'.
(BiGer6-10, 6;8 – A2)

Additionally, a few personal pronouns with a relative clause were found, e.g. *han som jobbade där* 'he who worked there' (MoSwe5-19, 5;4, A2). Both these types occurred almost exclusively to introduce the lifeguard in the A2 story.

Figure 6.3 presents the results for the proportion of pronouns by age and language groups for the ENNI narratives.

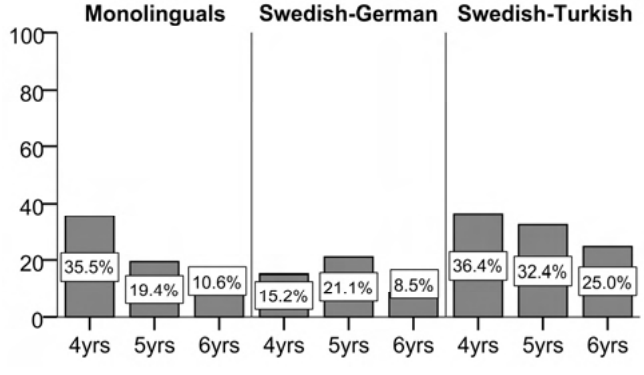

Figure 6.3. Character introductions, proportion (%) of pronouns by age and language group, Swedish ENNI (A2/B2).

A number of observation can be made from Figure 6.3. First, all groups produced a higher proportion of pronouns in their ENNI narratives than they did in MAIN1. In fact, the six-year-olds used similar, or in the case of the Swedish-Turkish bilinguals, even higher, proportions of pronouns in ENNI than the four-year-olds did in MAIN1 (recall Figure 6.2).

Second, the patterns in the three language groups were not identical. In the monolingual group, we see a relatively steep decline in the proportion of pronouns with age, i.e. substantial improvements both from age 4 to 5 and from age 5 to 6. The Swedish-German four-year-olds produced a lower proportion of pronouns than the five-year-olds, but the six-year-olds had the lowest proportion. In the Swedish-Turkish groups results were similar to those of the monolinguals in MAIN1 with equivalent proportions of pronouns produced by the two younger groups, but a lower proportion for the six-year-olds. Whereas the Swedish-Turkish and the monolingual four-year-olds produced similar proportions of pronouns, the older Swedish-Turkish groups did not perform as well as their monolingual peers, i.e. the development for the Swedish-Turkish children seems to be slower than for the monolinguals.

Next, results from the statistical analysis of pronouns versus lexical NPs are shown. In Table 6.8, the summary of the final logistic regression model of pronoun vs lexical NP for the Swedish data is shown.

Table 6.8. Summary of logistic regression model 6.1: Character introduction, pronoun versus lexical NP, Swedish MAIN1 (Cat/Dog) and ENNI (A2/B2).

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	-1.312	0.127	107.482	< .001***
Narrative task: MAIN1 vs ENNI	-1.249	0.210	35.188	< .001***
Age (1): 6 vs 4 & 5	-0.914	0.234	15.211	< .001***
Age (2): 5 vs 4	-0.400	0.226	3.134	.077
Language group (1): monolinguals vs bilinguals	-0.033	0.202	0.027	.868
Language group (2): Swedish-German vs Swedish-Turkish	-0.913	0.277	10.897	.001**
Model evaluation				
R^2 (Nagelkerke)	.129			
-2 Log likelihood	678.082			

Note. ** = $p < .01$, *** = $p < .001$. The second level of each predictor is the reference level.

The model in Table 6.8 shows a main effect of narrative task, indicating a significantly lower proportion of pronouns in MAIN1 compared with ENNI ($p < .001$). There was also a main effect of age for the six-year-olds compared with the younger groups ($p < .001$), i.e. fewer pronouns in the six-year-old group, whilst the difference between the younger groups did not reach significance ($p = .077$). No significant differences were found between monolinguals and bilinguals ($p = .868$), but the difference between the Swedish-Turkish and the Swedish-German bilinguals was highly significant ($p = .001$). No interactions were included in the model, because they did not contribute significantly to the model fit.

Although representing a logical division of the data into monolinguals versus bilinguals, the comparisons between the language groups in the model

shown in Table 6.8 do not give the full picture of the results seen above, since not all group comparisons are included. The monolinguals and the Swedish-German bilinguals performed similarly whereas results for the Swedish-Turkish bilinguals were different (cf. Figures 6.2 and 6.3). Therefore, a second model was run in which the variable language group was recoded so that one binary predictor compared the Swedish-Turkish bilinguals with the Swedish-German bilinguals and the monolinguals, and the second predictor compared the Swedish-German bilinguals and the monolinguals. The full regression model with this alternative coding of the variable language group can be found in Appendix 4, Table A4.5. The new version of the model for pronoun versus lexical NP showed that the Swedish-Turkish bilinguals produced a higher proportion of pronouns than the Swedish-German bilinguals and the monolinguals (β = 0.702, *SE* = 0.214, z = 10.788, $p < .001$), but that there was no difference in pronoun use between the Swedish-German bilinguals and the monolinguals (β = -0.423, *SE* = 0.265, z = 2.561, p = .110).

6.4.1.3 Fully appropriate NPs

Figure 6.4 shows the proportion (%) of fully appropriate NPs used by the different groups in MAIN1 (Cat/Dog).

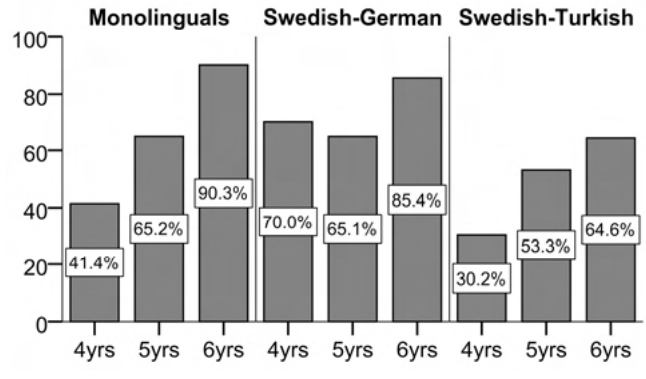

Figure 6.4. Character introductions, proportion (%) of fully appropriate NPs, by age and language group, Swedish MAIN1 (Cat/Dog). Fully appropriate NPs include indefinite NPs, possessive NPs and proper nouns.

In all three language groups, clear age effects can be seen, with the older groups producing higher proportions of fully appropriate NPs. The only exception is found in the Swedish-German bilingual group, in which the four- and five-year-olds produced similar proportions of fully appropriate NPs. The proportions of fully appropriate NPs produced by the monolingual and Swedish-German six-year-olds were comparable and very high, whereas the proportion in the Swedish-Turkish six-year-old group was substantially low-

er. In fact, the Swedish-Turkish six-year-olds performed at the same level as the monolingual and Swedish-German five-year-olds.

In Figure 6.5, proportions (%) of fully appropriate NPs are shown for the different age groups within each language group for the ENNI narratives.

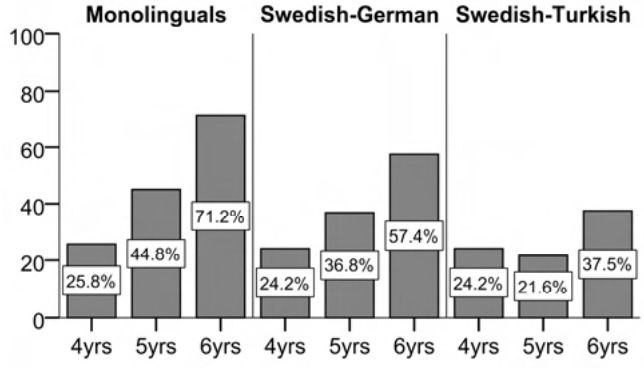

Figure 6.5. Character introductions, proportion (%) of fully appropriate NPs, by age and language group, Swedish ENNI (A2/B2). Fully appropriate NPs include indefinite NPs, possessive NPs and proper nouns.

In Figure 6.5, we see that the four-year-old groups used similar proportions of fully appropriate NPs in all language groups. For the older groups, there are differences between the language groups, but in all three language groups, the six-year-olds performed substantially better than the younger groups. The difference between five- and six-year-olds was largest in the monolingual group, slightly smaller in the Swedish-German group, and the smallest in the Swedish-Turkish group. In fact, the proportion of fully appropriate NPs used by the monolingual six-year-olds was almost twice as high as for the Swedish-Turkish bilinguals and almost 14 percentage points higher than the Swedish-German six-year-olds. In ENNI, similarly to in MAIN1, the Swedish-Turkish six-year-olds have a performance that is comparable to the five-year-olds in the other two language groups.

When comparing Figures 6.4 and 6.5, the amount of fully appropriate NPs is substantially lower in ENNI than in MAIN. The summary of the final logistic regression model for fully appropriate NP versus other NP, shown in Table 6.9, confirms this impression. There was a significant main effect of narrative task ($p < .001$); more fully appropriate NPs were used in MAIN1 than in ENNI. There were also highly significant effects of age: the six-year-olds produced more fully appropriate expressions than the younger groups ($p < .001$), and the five-year-olds also performed better than the four-year-olds ($p = .003$). Additionally, the monolinguals performed better than the bilinguals ($p = .003$) and the Swedish-German bilinguals used higher proportions of fully appropriate NPs than the Swedish-Turkish bilinguals ($p < .001$). There was also an interaction effect between age and language group. Only

one of the interaction predictors was significant, the interaction between monolinguals versus bilinguals and six-year-olds versus the younger groups (p = .009). Running separate Chi-square-tests on the proportions of fully appropriate NPs for the six-year-olds and the younger groups confirmed that there was a significant (and relatively large) difference between the monolinguals and the bilinguals in the proportion of fully appropriate NPs in the six-year-old group ($\chi^2(1, N = 321) = 13.437$, $p < .001$), but not in the younger groups ($\chi^2(1, N = 577) = 0.377$, $p = .539$). The interaction effect is shown in Figure 6.6.

Table 6.9. Summary of logistic regression model 6.2: Character introduction, fully appropriate NP vs other NP, Swedish MAIN1 (Cat/Dog) and ENNI (A2/B2).

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	-0.514	0.109	22.043	< .001***
Narrative task: MAIN1 vs ENNI	1.106	0.150	54.177	< .001***
Age (1): 6 vs 4 & 5	1.222	0.162	57.219	< .001***
Age (2): 5 vs 4	0.552	0.185	8.938	.003**
Language group (1): monolinguals vs bilinguals	0.443	0.151	8.654	.003**
Language group (2): Swedish-German vs Swedish-Turkish	0.846	0.195	18.741	< .001***
Age (1) x Language group (1)	0.860	0.328	6.869	.009**
Age (1) x Language group (2)	0.139	0.404	0.119	.730
Age (2) x Language group (1)	0.610	0.358	2.908	.088
Age (2) x Language group (2)	-0.439	0.486	0.814	.367
Model evaluation				
R^2 (Nagelkerke)	.224			
-2 Log likelihood	1076.837			

Note. ** = p < .01, *** = p < .001. The second value of each predictor is the reference level for that predictor.

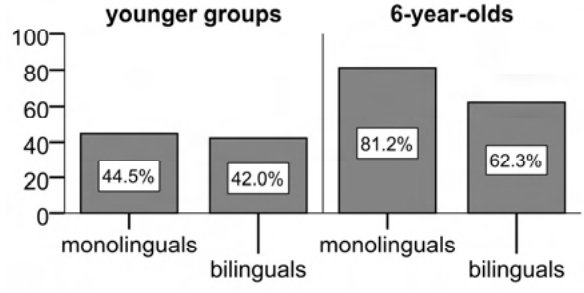

Figure 6.6. Fully appropriate lexical NPs, monolinguals vs bilinguals and six-year-olds vs the younger groups (four- and five-year-olds), Swedish (MAIN1 and ENNI combined).

Following the same procedure as in the analysis of pronouns versus lexical NPs, the variable language group was recoded (with one binary predictor comparing the Swedish-Turkish bilinguals with the Swedish-German bilin-

guals and the monolinguals, and a second predictor comparing the Swedish-German bilinguals and the monolinguals) and the model was then rerun. The full model with this alternative coding of language group can be found in Appendix 4, Table A4.6. The results showed a clear main effect for the Swedish-Turkish bilinguals versus the other groups (β = -0.856, SE = 0.166, z = 26.588, p < .001), but no difference between the monolinguals and the Swedish-German bilinguals (β = -0.020, SE = 0.178, z = 0.013, p = .910). The Swedish-Turkish bilinguals thus produced a significantly lower proportion of fully appropriate NPs than the other groups. Additionally, just as in the original model, there was an interaction effect between age and language group. The only significant interaction predictor was the interaction between the six-year-olds versus the younger groups and the monolinguals versus the Swedish-German bilinguals (β = -0.790, SE = 0.387, z = 4.167, p = .041). The significant interaction indicates that the relationship between the results for monolinguals and the Swedish-German bilinguals is not the same for the six-year-olds as for the younger groups combined. However, when separate Chi-square tests were run for the six-year-olds and the younger groups, there was no significant difference between monolinguals and Swedish-German bilinguals in either group (six-year-olds: $\chi^2(1, N = 233) = 2.939$, p = .086; younger groups: $\chi^2(1, N = 419) = 1.466$, p = .226).

To summarize the results for character introduction in Swedish, there were clear effects of age, language group and narrative task. The six-year-olds produced a lower proportion of pronouns and a higher proportion of fully appropriate referring expressions than both younger groups. There were no differences between the two younger groups in proportions of pronouns, but the five-year-olds produced a higher proportion of fully appropriate expressions than the four-year-olds. The Swedish-Turkish bilinguals produced a higher proportion of pronouns and a lower proportion of fully appropriate expressions than the monolinguals and Swedish-German bilinguals, who performed similarly. Finally, the children performed better, with fewer pronouns and more fully appropriate expressions, in MAIN Cat/Dog than in ENNI A2/B2.

6.4.2 Swedish-German bilinguals

In this section, results from the Swedish and German narratives are compared for MAIN1 and ENNI and for the three age groups of the Swedish-German bilinguals. Table 6.10 gives an overview of the number of referring expressions produced by the Swedish-German bilinguals in Swedish and German for MAIN1 and ENNI, respectively.

Table 6.10. Number of referring expressions used for introducing story characters in MAIN1 (Cat/Dog) and ENNI (A2/B2), Swedish-German bilinguals (N=46), by language.

	MAIN1	ENNI	Total
Swedish	131	118	249
German	131	117	248
Total	**262**	**235**	**497**

In total, 497 referring expressions were produced by the Swedish-German bilinguals, with an equal number in each language. A lower number of expressions was produced in ENNI, partly because the number of children telling an ENNI narrative was lower, and partly because of the use of one referring expression to simultaneously introduce two characters. The mean number of characters introduced in the German ENNI was 2.8 (median: 3), which was identical to the number of characters introduced in the German MAIN1. The mean number of referring expressions used was 2.9 for MAIN1 and 2.7 for ENNI.[103]

6.4.2.1 Types of referring expressions

There was no difference between the first and the second testing in the distribution of different types of referring expressions ($\chi^2(5, N = 497) = 5.979$, $p = .308$). The variable order of testing was therefore not included in further analyses.

Table 6.11 gives an overview of the types of referring expressions found in the MAIN1 narratives for Swedish and German for all age groups together.[104] No significant differences were found in the distribution of types of referring expressions between the German Cat and Dog ($\chi^2(3, N = 262) = 5.411$, $p = .144$), and the data from these two stories were therefore combined for all analyses (for a comparison of the Swedish MAIN1 data, see Section 6.4.1.1).

In both languages, the majority of the referring expressions used by the children were indefinite NPs, but these were somewhat more common in Swedish than in German. Proportions of definite NPs were slightly higher in German. Proportions of pronouns and bare nouns were low in both languages.

[103] For details about the number of referring expressions per child (excluding the mixed cases found in the German data, described in 6.3), as well as the number of referring expressions in the German data divided by age group, see Appendix 4, Table A4.7-A4.8.

[104] For detailed overviews of all types of referring expressions (including raw figures) in the German narratives by age group, see Appendix 4, Table A4.9. The same results for Swedish are found in Appendix 4, Table A4.3-A4.4.

Table 6.11. Types of referring expressions used to introduce story characters, percentages (%), MAIN1 (Cat/Dog) by language, Swedish-German bilinguals (all age groups).

	Swedish (N=131)	**German (N=131)**
Pronouns	5.3	6.9
Bare Nouns	4.6	2.3
Definite NPs	16.0	25.2
Indefinite NPs	74.0	65.6
Total	**100.0**	**100.0**

Note. N = total number of referring expressions

No difference was found in the German ENNI data between the two stories (A2 vs B2) in the distribution of different types of referring expressions ($\chi^2(4, N = 235) = 6.566$, $p = .161$); the lack of a difference between the stories in the Swedish data has been shown above (Section 6.4.1.1). The types of referring expressions found in ENNI are shown in Table 6.12, by language.

Table 6.12. Types of referring expressions used to introduce story characters, percentages (%) ENNI (A2/B2) by language, Swedish-German bilinguals (all age groups).

	Swedish (N=118)	**German (N=117)**
Pronouns	14.4	11.1
Bare Nouns	4.2	3.4
Definite NPs	39.8	33.3
Indefinite NPs	36.4	51.3
Possessive NPs	3.4	0.9
Proper Nouns	1.7	0.0
Total	**100.0**	**100.0**

Note. N = total number of referring expressions.

In MAIN1, the children seemed to perform somewhat better in Swedish than in German. The opposite trend can be seen in the ENNI data. In the German ENNI data, the proportions of pronouns and definite NPs were somewhat lower and the proportion of indefinite NPs higher than in the Swedish ENNI data. The children produced more possessive NPs and proper nouns in Swedish, although these proportions were generally low. Similar to the Swedish ENNI results, some plural pronouns were found in the German data, as in (20).

(20) <u>die beiden</u> wollen schwimmen gehen
'<u>They both</u> want to go swimming.' (BiGer5-10, 5;0 – A2).

Figure 6.7 gives an overview of the difference between MAIN1 and ENNI in the two languages for types of referring expressions with a focus on the pro-

portions of pronouns and fully appropriate NPs. In both languages, proportions of pronouns are higher and proportions of fully appropriate NPs are lower in ENNI than in MAIN.

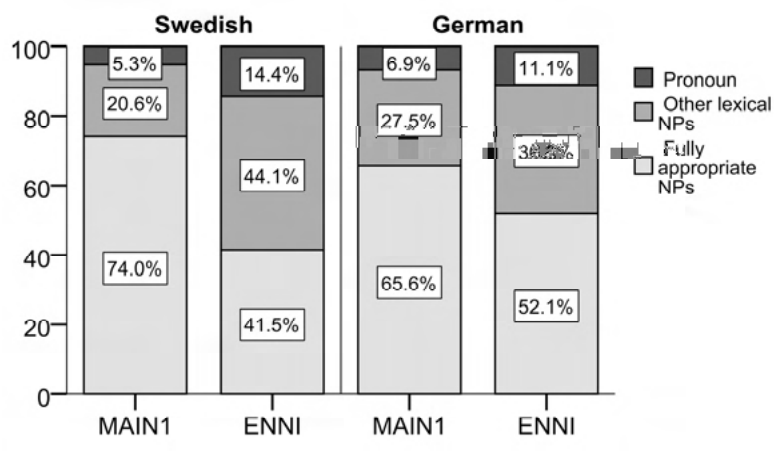

Figure 6.7. Character introductions, proportions (%) of fully appropriate NPs, other lexical NPs, and pronouns by language and narrative task, MAIN1 (Cat/Dog) and ENNI (A2/B2), Swedish-German bilinguals.

6.4.2.2 Pronouns

Figure 6.8 shows the proportions of pronouns in both languages for the different age groups in the MAIN1 narratives. The five- and six-year-olds produced similar low proportions of pronouns in both German and Swedish. The four-year-olds used a higher proportion of pronouns than the older groups, and the proportion was higher for German than for Swedish. In Swedish, almost all pronouns used were personal pronouns. In German, demonstrative pronouns were the most common type.

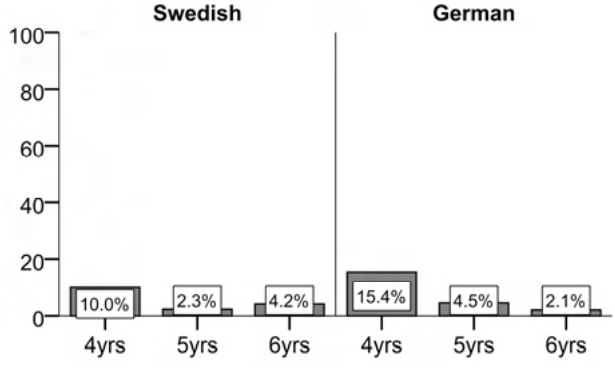

Figure 6.8. Character introductions, proportion (%) of pronouns by age group and language, Swedish-German bilinguals, Swedish and German MAIN1 (Cat/Dog).

In Figure 6.9, the proportions of pronouns in the Swedish and German ENNI data are shown for the three age groups.

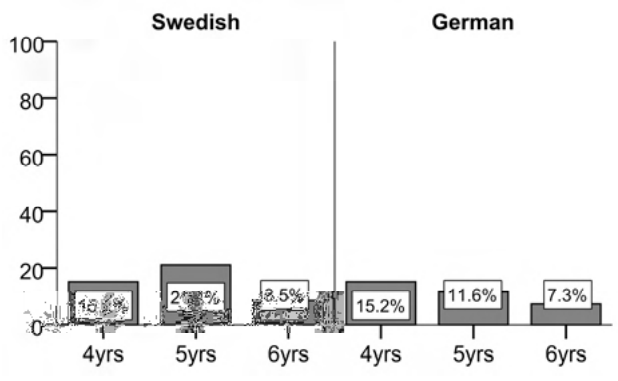

Figure 6.9. Character introductions, proportion (%) of pronouns by age group and language, Swedish-German bilinguals, Swedish and German ENNI (A2/B2).

Proportions of pronouns were generally higher in ENNI than in MAIN1, with the exception of the four-year-olds in German, where pronoun proportions were comparable (cf. Figure 6.8). When ENNI-results from Swedish and German are compared, we see some differences between the age groups. The four-year-olds produced identical proportions of pronouns in both languages and the proportions produced by the six-year-olds were very similar. The five-year-olds produced almost twice the proportion of pronouns in Swedish than in German. The four-year-olds performed better than the five-year-olds in Swedish, but not in German. The six-year-olds performed better than the younger children in both languages.

Table 6.13 shows the summary of the final logistic regression model for pronoun versus lexical NP for the Swedish-German bilinguals.

Table 6.13. Summary of logistic regression model 6.3: Character introduction, pronoun versus lexical NP, MAIN1 (Cat/Dog) and ENNI (A2/B2), Swedish-German bilinguals.

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	-1.996	0.259	59.577	< .001***
Narrative task: MAIN1 vs ENNI	-0.839	0.326	6.605	.01*
Age (1): 6 vs 4 & 5	-0.837	0.373	5.028	.025*
Age (2): 5 vs 4	-0.446	0.360	1.536	.215
Language: Swedish vs German	0.107	0.315	0.116	.733
Model evaluation				
R^2 (Nagelkerke)	.06			
-2 Log likelihood	292.654			

Note. * = p < .05, *** = p < .001. The second value of each predictor is the reference level for that predictor.

There was a clear main effect of narrative task (p = .01). Significantly fewer pronouns were used in MAIN1 than in ENNI. There was also a significant effect of age; the six-year-olds produced fewer pronouns than the younger groups (p = .025), but there was no difference between the two younger groups (p = .215). There was no significant effect of language (p = .733). No interactions were included in the model, because they did not contribute to the model fit.

6.4.2.3 Fully appropriate NPs

In Figure 6.10, the proportions of fully appropriate NPs in Swedish and German are shown for the three age groups, for MAIN1.

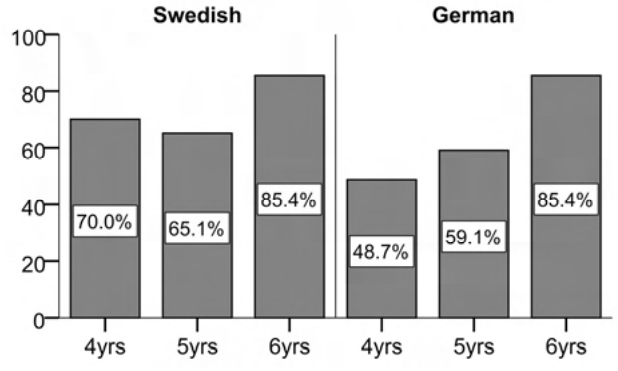

Figure 6.10. Character introductions, proportion (%) of fully appropriate NPs by age group and language, Swedish-German bilinguals, Swedish and German MAIN1 (Cat/Dog). Fully appropriate NPs include indefinite NPs, possessive NPs and proper nouns.

As has already been shown in Section 6.4.1.3, in Swedish, performance was high already at age 4, and there was a relatively large difference between the two younger groups (who performed similarly) and the six-year-olds. The German pattern is somewhat different. Whereas the six-year-olds performed identically in the two languages, the younger groups had lower performance in German. The difference between the languages was large for the four-year-olds, but relatively small for the five-year-olds. It should also be noted that the Swedish-German four-year-old bilinguals' result for German is still higher than the monolingual Swedish four-year-old's results (41.4% fully appropriate NPs for the monolingual Swedish four-year-olds in MAIN1, cf. Figure 6.4).

Figure 6.11 shows the proportions of fully appropriate NPs in the Swedish and German ENNI narratives by age group. The six-year-olds produced similar proportions of fully appropriate NPs in both languages, but both younger groups performed better in German. The difference between the languages was especially pronounced in the four-year-olds, but the differ-

ence between the languages for the five-year-old group was also relatively large. There thus seems to be a clearer age development in Swedish.

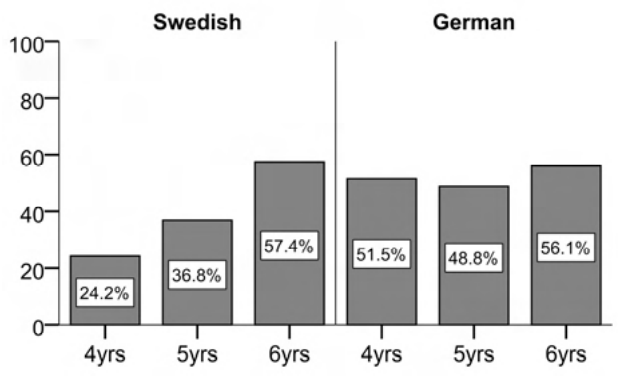

Figure 6.11. Character introductions, proportion (%) of fully appropriate NPs by age group and language, Swedish-German bilinguals, Swedish and German ENNI (A2/B2). Fully appropriate NPs include indefinite NPs, possessive NPs and proper nouns.

In Table 6.14, the final logistic regression model of fully appropriate NP versus other NP for the Swedish-German bilinguals is shown.

Table 6.14. Summary of logistic regression model 6.4: Fully appropriate NP versus other NP, MAIN1 (Cat/Dog) and ENNI (A2/B2), Swedish-German bilinguals.

Predictor	***β***	**SE**	***z* (Wald)**	***p* value**
Intercept/Constant	0.071	0.190	0.139	.709
Narrative task: MAIN1 vs ENNI	0.579	0.268	4.682	.03*
Age: 6 vs 4 & 5	0.969	0.208	21.678	< .001***
Age: 5 vs 4	0.168	0.236	0.506	.477
Language: Swedish vs German	-0.496	0.271	3.354	.067
Language x Narrative task	0.915	0.388	5.569	.018*
Model evaluation				
R^2 (Nagelkerke)	.142			
-2 Log likelihood	617.545			

Note. * = p < .05, *** = p < .001

There is a significant main effect of narrative task (p = .03); more fully appropriate expressions were produced in MAIN1 than in ENNI. Additionally, the six-year-olds produced a higher proportion of fully appropriate NPs than the younger groups (p < .001). There was no significant difference between the two younger groups (p = .477). The main effect of language did not reach significance (p = .067), but there was a significant interaction between language and narrative task (p = .018). This interaction effect requires some further explanation. Although there was a significant effect of narrative task in both languages (Swedish: $\chi^2(1, N = 249) = 27.069$, p < .001; German: $\chi^2(1, N = 248) = 4.674$, p = .031), the effect of narrative task was stronger in

Swedish, i.e. the difference between MAIN1 and ENNI was larger in Swedish. This interaction effect can be seen in Figure 6.7 above (in light grey); the difference between MAIN1 and ENNI was 32.5 percentage points in Swedish (74% vs 41.5%), but only 13.5 percentage points in German (65.6% vs 52.1%). In Swedish, the Swedish-German bilinguals thus produced a higher proportion of fully appropriate NPs in MAIN1, but a lower proportion in ENNI, compared to their performance in German.

To summarize the results from Swedish-German bilinguals' character introductions, there were effects of age, with the six-year-olds performing better than the younger groups, and narrative task, with better performance in MAIN than in ENNI. The difference between MAIN and ENNI was larger in Swedish for fully appropriate NPs.

6.5 Summary of results and discussion

In this chapter, the use of referring expressions to introduce story characters in the MAIN1 (Cat/Dog) and ENNI (A2/B2) narratives has been investigated. A total of 325 Swedish narratives from all three language groups and 90 German narratives from the Swedish-German bilinguals were analyzed. All introductions of story characters were classified according to type of referring expression (e.g. pronoun, indefinite NP). The analysis focused on use of pronouns and fully appropriate NPs, respectively. The following research questions were asked: (1) Is there a development with age in the use of appropriate referring expression for introducing story character? (2) For Swedish, are there any differences between the language groups? (3) Do the Swedish-German bilinguals perform differently in their two languages? (4) Is there any effect of the stimulus material, i.e. is there a difference in performance between MAIN1 (Cat/Dog) and ENNI (A2/B2)?

The results give clear answers to all four questions. First, results from character introduction in the Swedish data from the MAIN1 and ENNI showed significant effects of age; the six-year-olds produced fewer pronouns and more fully appropriate NPs than the younger groups. There was no difference between the two younger groups in the proportions of pronouns, but the five-year-olds did produce significantly higher proportions of fully appropriate NPs than the four-year-olds. For both languages of the Swedish-German bilinguals, there were differences between the six-year-olds and the younger groups; the four- and five-year-olds performed similarly. Second, there was a significant effect of language group: the Swedish-Turkish bilinguals produced higher proportions of pronouns and lower proportions of fully appropriate NPs than the other two groups. There were no significant differences between the Swedish-German bilinguals and the Swedish monolinguals. Third, there were no significant differences between the Swedish-German bilinguals' two languages, except that the difference between

MAIN1 and ENNI in the proportion of fully appropriate NPs was larger in Swedish than in German. Fourth and finally, for Swedish and for both languages of the Swedish-German bilinguals, there were clear effects of the stimulus material (narrative task) on both the proportion of pronouns and the proportion of fully appropriate NPs; the children used fewer pronouns and introduced characters appropriately to a larger extent in their MAIN1 narratives compared with their ENNI narratives. In the following, each of these results will be discussed in turn, starting with the age effects.

The children's overall performance is high, possibly with the exception of the youngest children on ENNI. All Swedish-speaking four- to six-year-olds in the current study, mono- as well as bilinguals, performed well compared to results from earlier studies of monolingual children speaking other languages, including Turkish and German (e.g. Hickmann et al., 1996; Kail & Hickmann, 1992; Kail & Sanchez y Lopez, 1997; Küntay, 2002), as has already been shown for the Swedish monolinguals in Lindgren (2018). The ability to introduce characters appropriately in narrative discourse thus seems to develop relatively early in Swedish-speaking children. Additionally, the Swedish-German children showed a substantially higher proportion of fully appropriate NPs in both languages than the Russian-German bilinguals did in German in the studies by Topaj (2010) and Reichardt (2014). The Swedish-German bilinguals of the current study also performed above the German monolinguals in Hickmann et al. (1996). However, there was also substantial development from age four to six, consistent with earlier studies (e.g. Aksu-Koç & Nicolopoulou, 2015; Schneider & Hayward, 2010). Although the four-year-olds in the current study performed well in that they did use substantial proportions of fully appropriate NPs, especially in their MAIN1 narratives, they were far from consistent in their choice of referring expressions; they chose less appropriate expressions, such as definite NPs and pronouns more often than they chose fully appropriate expressions, such as indefinite NPs. Swedish-speaking four-year-olds have thus not mastered the correct use of referring expressions when introducing story characters in these two narrative tasks, MAIN Cat/Dog and ENNI A2/B2. The performance of the six-year-olds was substantially higher. In fact, the monolingual and Swedish-German six-year-olds almost exclusively used fully appropriate NPs to introduce characters in the MAIN1 narratives. The five-year-olds fell in between the four- and the six-year-olds, using a majority of appropriate expressions, but not to the same extent as the six-year-olds. In line with earlier studies, this indicates that the late preschool years, e.g. age 4–6 are central for the development of the ability to introduce referents appropriately (cf. Aksu-Koç & Nicolopoulou, 2015).

Moving on to the effect of language group, the results suggest that the Swedish-Turkish children master appropriate introductions of story characters in Swedish somewhat later than Swedish-German bilinguals and Swedish monolinguals, at least when this ability is measured in picture-based

narrative tasks. As the Swedish-German bilinguals performed similarly to the monolinguals, the results from the Swedish-Turkish children cannot be attributed to a general effect of bilingualism. There are a number of other possible explanations for the lower performance of the Swedish-Turkish children. First, it could be linked to *differences in the referential systems* of their two languages Swedish and Turkish (cf. Section 6.2). Such differences may lead to cross-linguistic influence in the children's productions, with transfer of forms from the first or the stronger language to the second or weaker language, i.e. from Turkish to Swedish. Influence from Turkish would manifest itself in a high(er) percentage of bare nouns, as these are the preferred form for introducing referents. In MAIN1, the Swedish-Turkish bilinguals produced equal proportions of bare nouns compared with the children in the other language groups and, in ENNI, only the younger Swedish-Turkish children produced a higher proportion of bare nouns than the children in the other language groups. These results seem to support the transfer explanation at least at the earlier stages of the development and for the more difficult narrative stimulus only. A closer analysis revealed that the Swedish-Turkish bilinguals' many bare nouns in the ENNI narratives were mainly used to introduce a specific character, namely the doctor in the ENNI B2-story.[105] There thus seems to be transfer in this specific case, but not for character introductions in general. The current study does thus not show strong support for cross-linguistic influence from Turkish as the main explanation for the Swedish-Turkish bilinguals' performance in Swedish.

There is another plausible explanation linked to the differences between the referential systems. Compared to learning Swedish and German, where referential systems are virtually identical (differing only in the type of definiteness marker), learning two languages with different systems, such as Swedish and Turkish, may be more demanding for the child. This could ultimately lead to a mastery of appropriate use of referring expressions in Swedish. This is in line with findings from earlier studies that children speaking two languages with different referential systems may master appropriate referent introduction later than monolinguals (Chen & Lei, 2013; Chen & Pan, 2009; Jia & Paradis, 2015; Reichardt, 2014; Topaj, 2010) whereas children acquiring two languages with similar referential systems perform similarly to monolinguals (Finnstedt, 2013; Serratrice, 2007). This could also explain why the Swedish-German bilinguals performed similarly in their two languages; when learning to introduce characters in their weaker language German, they would be helped by having a similar referential system in their stronger language Swedish. The fact that the Swedish-German four-year-olds performed even better than their monolingual Swedish peers could be explained by their performance being 'boosted' by the similarities

[105] Remember that this noun is always used in bare form in Turkish.

of the system, leading to faster acquisition of the use of appropriate referring expressions in this type of bilinguals.

There are also indications in the literature that monolingual Turkish children acquire appropriate use of referring expressions for character introduction later than children speaking other languages, as was found in comparisons with English and Greek (Aksu-Koç & Nicolopoulou, 2015). If the ability to introduce referents appropriately develops later (or slower) in Turkish-speaking children than for example in Swedish-speaking children, it may also contribute to a slower pace of acquisition for Swedish-Turkish bilinguals compared with other Swedish-speaking bilinguals. In order to investigate this issue further, it would be necessary to look both at the Swedish-Turkish children's Turkish productions, as well as data from Turkish monolinguals on similar tasks as in the current study.

Another factor that may have an impact on the ability to introduce referents appropriately in this type of narrative task is *language proficiency*. The lower performance of the Swedish-Turkish bilinguals could be linked to their overall lower Swedish proficiency, as rated by their parents (cf. Section 3.1.2.3) and shown in their lower lexical knowledge (cf. Section 5.2.2), as well as in the overall level of grammatical correctness in their narratives (as judged impressionistically). This could mean that they have acquired the morphological means to correctly introduce referents to a lesser degree than the monolinguals and Swedish-German bilinguals, e.g. as regards the acquisition of indefinite articles. From the analyses performed within the current study, we cannot determine whether a child has fully acquired the indefinite articles. However, 62.5% (10/16) of the Swedish-Turkish four-year-olds, as well as a few five- and six-year-olds in this language group did not use *any* indefinite NP when introducing the character in MAIN1 (which was the narrative task in which their performance was higher). It could be the case that these children have not yet acquired the indefiniteness markers. As a comparison, one third (8/24) of the monolingual four-year-olds did not use any indefinite NPs in their MAIN1 character introductions. As the indefinite articles are typically acquired around or even before age two in monolingual Swedish (Bohnacker, 2003, 2007), it seems unlikely that the monolingual four-year-olds in the current study have not yet acquired the indefinite articles. It is more likely that these children, both mono- and bilingual, have not yet learned to *use* indefinite NPs to introduce referents in this type of narrative task. This may of course also be linked to incomplete acquisition of the ability to take the listener's perspective into account.

The last point to discuss is the effects of the stimulus material (narrative task) on the children's performance. In the current study, the stimulus material affected the different groups in similar ways, with all groups performing lower in ENNI A2/B2 than on MAIN1 Cat/Dog. Results were the same for Swedish and for both languages of the Swedish-German bilinguals. Properties of the stimulus materials may influence how demanding the task is,

which in turn influences child performance. This is especially true in the current study, where procedures were more or less identical for MAIN1 Cat/Dog and ENNI A2/B2 (cf. Sections 3.2.2 and 3.2.3), but where the stimulus materials, i.e. the pictures, are very different. Despite containing the same number of characters, and being of similar length, MAIN Cat/Dog and ENNI A2/B2 differ in a number of important ways. For example, there is a clear difference in the type of characters (Cat/Dog contain two animals and one human character, whereas A2/B2 have three humanized animal agents), and the point at which characters enter the story (cf. Section 3.2.3, Table 3.11), and in the number of main characters. The MAIN-pictures are in full color whereas the ENNI-pictures are black-and-white line drawings. All these aspects may contribute to the ENNI being a more demanding task for the child which makes it more difficult to consistently produce fully appropriate referring expressions. For a more detailed discussion of the differences between these two elicitation instruments, see Lindgren (2018). The current study has shown that the choice of stimulus material may affect the conclusions drawn regarding the age at which appropriate referring expressions are used consistently to introduce story characters.

Some final words concerning the performance of the groups are warranted. For character introduction, the Swedish-Turkish six-year-olds were roughly one year behind their monolingual peers. The Swedish-Turkish four-year-olds performed relatively similarly to their monolingual and Swedish-German age peers, indicating that it is not the case that development starts later in Swedish-Turkish bilinguals, but rather that the process of acquisition is somewhat slower. It remains an open question at which point the Swedish-Turkish children will catch up with their monolingual peers.

To conclude, there were clear effects of age, language group, and task on character introductions in the children's narratives. The six-year-olds outperformed both younger groups. The four- and five-year-olds performed similarly in terms of (inappropriate) pronouns, but the five-year-olds produced more fully appropriate NPs. The Swedish-Turkish bilinguals performed lower than the Swedish-German bilinguals and the monolinguals. The Swedish-German bilinguals performed equally well in Swedish and German. Results were highly task-dependent. Performance was higher on MAIN Cat/Dog than on ENNI A2/B2. The age at which children master the use of appropriate referring expressions for introducing story characters thus appears to vary depending on the task.

7 Narrative macrostructure

Narrative macrostructure is the most central aspect of narrative competence, as it concerns the narrative's global structure (cf. Chapter 1). This is of course linked to which story content is expressed by the child, and without content there is no narrative. Generally, comprehension of underlying narrative schemata (cf. Section 1.1) is taken as a prerequisite for being able to produce narrative content in a well-formed manner (Shapiro & Hudson, 1991; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994), yet comprehension itself, for example in the form of answers to probe questions which require inferencing on the part of the child, has not often been investigated (cf. Section 7.1). Also, little is known about Swedish-speaking children's narrative comprehension and production. The aim of this chapter is therefore to investigate both comprehension and production of narrative macrostructure following the MAIN scoring protocol. The Swedish, German and Turkish MAIN narratives and scores for the accompanying comprehension questions are analyzed. The following specific questions are asked:

- How does *comprehension* of narrative macrostructure develop from age 4 to 6?
- How does *production* of narrative macrostructure develop from age 4 to 6?
- Do comprehension and production of narrative macrostructure in Swedish differ between the three language groups?
- Do the bilinguals perform differently in their two languages (Swedish vs German; Swedish vs Turkish)?
- Are there any differences in narrative macrostructure between the two minority languages German and Turkish?
- Are there differences between comprehension and production and between the two narrative tasks (MAIN1 Cat/Dog vs MAIN2 Baby Birds/Baby Goats)?

This chapter proceeds as follows. After summarizing insights from earlier studies of narrative macrostructure (Section 7.1), Section 7.2 reports results from comprehension of macrostructure. Section 7.3 deals with production of macrostructure. In Section 7.4, results from comprehension and production are compared. The chapter concludes with a discussion and summary of the results (Section 7.5).

7.1 Narrative macrostructure: Insights from earlier studies

There is relatively large variation between studies in what is included in macrostructure, both in the type of stimulus material and the specific story grammar models used for data analysis (cf. Chapter 1). This makes it difficult to generalize the results of earlier studies. However, there are two common findings: (1) Narrative macrostructure develops extensively throughout the preschool and early school years (e.g. Berman & Slobin, 1994b; Mäkinen, 2014; Schneider et al., 2006; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Trabasso et al., 1992) and (2) bilinguals tend to perform similarly on macrostructure in their two languages (e.g. Bohnacker, 2016; Fiestas & Peña, 2004; Iluz-Cohen & Walters, 2012; Pearson, 2002; Uccelli & Páez, 2007). This means "that the macrostructural narrative skills of children appear to vary with and depend more on age than on the language a story is told in" (Bohnacker, 2016, p. 22). In this section, broad trends regarding mono- and bilingual children's development of macrostructure in elicited oral picture-based fictional narratives are summarized first (Section 7.1.1), with one subsection on monolingual children (Section 7.1.1.1) and one on bilinguals (Section 7.1.1.2), followed by a section summarizing recent studies using the MAIN protocol for scoring macrostructure (Section 7.1.2). Only results from studies using the telling mode (cf. Section 1.4.1) are included.

7.1.1 General findings

Before describing the general findings for studies of monolinguals (Section 7.1.1.1) and bilinguals (Section 7.1.1.2), something should be said about studies carried out in a Swedish context. Generally, little is known about the development of narrative macrostructure in Swedish-speaking children. With the exception of the studies that have used the MAIN with Swedish-speaking bilinguals reported below (Section 7.1.2), a few studies of Swedish monolinguals have been published. These studies used the Frog story (Nordqvist, 2001) or shorter (six-picture) adaptations of other books in the Frog story series (Reuterskiöld et al., 2011; Reuterskiöld Wagner et al., 1999), or analyzed retellings of the Bus story (Renfrew, 1969) in children with language delay (Miniscalco, Hagberg, Kadesjö, Westerlund, & Gillberg, 2007). Additionally, there are a number of unpublished MA theses on mono- and bilingual Swedish-speaking children (e.g. Dillström & Kesti, 2009; Frithiofsson & Öberg, 2012; Nilsson & Vikström, 1997).

Earlier studies have used a wide range of different stimulus materials, many of which were originally not created for research, e.g. the Frog story (Mayer, 1969), which is a wordless picture book. When investigating both monolingual and bilingual (Swedish-speaking) children, it is warranted to

use an instrument that was specifically developed for narrative assessment and to use a standardized procedure for coding macrostructure, as is done in the current study.

7.1.1.1 Macrostructure in monolinguals

A number of studies of monolinguals, including some that used the Frog story as stimulus material, have not analyzed macrostructural components or story elements, but rather focused on narrative content more broadly, e.g. in the terms of information units or number of propositions expressing relevant narrative content (Hudson & Shapiro, 1991; Mäkinen, 2014; Mandler & Johnson, 1977; Norbury & Bishop, 2003; Reilly et al., 2004; Viberg, 2001). As this type of analysis does not concern macrostructure per se, they will not be discussed further here. With regard to age development, their results tend to be similar to those of studies analyzing macrostructural elements or story components.

In their seminal study of narratives elicited with the Frog story (Mayer, 1969) from monolingual children aged 3–4, 5, 7 and 9 and adults speaking English, German, Hebrew, Spanish and Turkish, Berman & Slobin (1994b) analyzed the overall plotline in terms of what they named *core plot components*. These components were based on the work of Labov & Waletsky (1967) (see also Labov, 1972, as well as Section 1.1), and consisted of the *onset* of the plot ('beginning'), *unfolding* of the plot ('middle') and *resolution* of the plot ('end'). Berman & Slobin (1994b, pp. 48–49) scored all narratives for the inclusion of these components. They found a clear development with age, with children aged 3 rarely including any of the three components, children aged 5 doing so around 50% of the time, and adults always including them. There were differences between the three components in the age at which they were included by most children – the resolution was only consistently mentioned by the adults, the unfolding was produced by almost all 9-year-olds, whereas the onset was already included by the majority of the 5-year-olds. At age 3–4, there was some variation between the languages, but the general patterns were consistent across all the five languages from age 5. Only a third of the five-year-olds included all three components, compared with two-thirds of the nine-year-olds, and almost all adults. The same type of analysis was used in Nordqvist (2001), one of the few studies in which the macrostructure of Swedish-speaking children's narratives are investigated. Nordqvist (2001) analyzed Frog story narratives told by three-, four-, five-, nine, and fifteen-year-old monolingual Swedish children and adults (N=84, 14 participants in each group). She found development with age, especially between age 4 and 5, for each of the three components (Nordqvist, 2001, pp. 201–203). Although the inclusion of onset, unfolding and resolution does show some age differences, it is a relatively coarse measure. In order to provide a more nuanced picture of the macrostructure, a

more fine-grained method of analysis needs to be used, such as one in which episodic (macrostructural) components are identified and scored.

Goals and so-called goal-attempt-outcome (GAO) sequences or 'full episodes' are central to many models for analyzing macrostructure, including the one MAIN is based on (cf. Section 1.1), as, in these models, narratives are seen as centered on protagonists' goals. Using a more detailed coding scheme for macrostructure of English Frog story narratives told by 3-, 5-, 9-year-olds and adults, Trabasso & Nickels (1992) found sharp increases in the use of GAO-sequences, i.e. full episodes, between age 3 and 5, and between age 5 and 9. Full episodes were very rarely produced by children at age 3. Further, Trabasso et al. (1992) found that although monolingual children who told the Frog story started to produce goals around age 5, not until age 9 did proportions of goals reach the same level as in adults. In contrast to the results for goals, Trabasso et al. (1992) found that attempts and outcomes were produced relatively frequently already at age 4.

In a study of 216 Estonian monolingual 6–7-year-olds, Soodla & Kikas (2010) found no effect of age on the inclusion of seven different story grammar elements (setting, initiating event, internal response, internal plan, attempt, consequence, reaction), or on the number of story information units, which is not surprising considering the small age range. Just as in the studies of Trabasso & Nickels (1992) and Trabasso et al. (1992), Soodla & Kikas (2010) found that some types of components were frequently expressed by the children in their narratives, whereas others were not. Initiating event, attempt and consequence were included by almost all children, whereas internal responses and reactions were only produced by a minority of children, and only 1 child out of 216 produced an internal plan (which could be seen as to some extent similar to the goal-category of the current study). Soodla & Kikas' (2010) results show that even children aged 6–7 may only infrequently include goals. It might also be the case that the frequency of goals depends on the stimulus material used, e.g. how salient the characters' goals are in the pictures.[106] Results from these earlier studies thus indicate that age development is not the same for different types of macrostructural components, and that it is important not only to analyze narratives in terms of an overall score for macrostructure but also to look more closely at different types of components. Therefore, the current study investigates different types of components in addition to a total macrostructure score.

Schneider et al. (2006) analyzed the inclusion of story grammar elements in narratives elicited with one simple (single-episode) and one more complex story (with three episodes) from the Edmonton Narrative Norms Instrument (ENNI). Three hundred English-monolingual typically-developing and 77

[106] There are indications that "for young children, the salience of individual pictured scenes is what counts, rather than a structurally motivated hierarchy of narrative importance" (Berman & Slobin, 1994b, p. 62).

children with LI aged 4–9 participated in the study and each child told both stories. Schneider et al. (2006) found that older children generally included a higher number of story grammar units and concluded that "the acquisition of story schema knowledge appears to develop as a function of age" (Schneider et al., 2006, p. 226). They also found differences between the simple and the more complex story, with scores for the simple story (ENNI A1, max score = 13) increasing until age 7 and then leveling off, whereas scores for the more complex story (A3, max score = 36) continued to increase until age 8. This shows that age effects are also dependent on the stimulus material: a simple story may not show the same differences between age groups as a more complex one, for example due to ceiling effects in scores.

Far fewer studies have investigated narrative comprehension than production. In fact, comprehension is not always clearly separated from production, and story production methods such as telling and retelling are sometimes listed as methods for investigating children's narrative comprehension (cf. Burris & Brown, 2014). A number of studies have investigated children's comprehension of narratives using probe questions (e.g. Stein & Glenn, 1979; Trabasso et al., 1992; see also the studies using MAIN in Section 7.1.2). Their results point to children's comprehension of narrative structure developing earlier than their ability to verbalize these structures in their narratives. In a classic study, Stein & Glenn (1979) analyzed answers to narrative comprehension questions (most of these targeted inferences) by 24 English monolingual six- and ten-year-olds. They found that 6-year-olds had overall good narrative comprehension, even of goals and internal states, story components which were rarely overtly expressed in the children's own narratives. Similarly, Trabasso et al. (1992) found that English monolingual four-year-olds did not spontaneously produce goals of story characters in their narratives but understood them, i.e. they could give reasons why a character performed an action when explicitly probed. These results show that "analyzing only what the speaker says may underestimate what the speaker knows" (Trabasso & Rodkin, 1994, p. 103). Thus, assessing children's narrative comprehension by eliciting productions (telling or retelling) of narratives may give a misleading picture of their actual understanding. Although not independent of general language proficiency, as answers still need to be expressed verbally, comprehension (probe) questions constitute a more direct way of measuring comprehension (cf. Liles, 1993).

7.1.1.2 Macrostructure in bilinguals

Results from studies of bilingual children point in similar directions as those of monolinguals with regards to age development in the production of narrative macrostructure. For example, in her study of Frog story narratives of 160 English-Spanish bilinguals and 80 English monolinguals in grade 2 (7-8-year-olds) and grade 5 (10-11-year-olds), Pearson (2002) found that at age 7–8, the bilinguals lagged behind on both story score (including what she

termed story elements) and a number of different language measures, including vocabulary. Although the bilinguals continued to lag behind in vocabulary and were less accurate in terms of morphosyntax, at age 10–11, they performed similarly to monolinguals on the story score. Story scores in the two languages were also significantly correlated in both age groups, although language scores were not, indicating that narrative ability is, at least to some extent, disconnected from language proficiency. There were clear differences between age 7–8 and age 10–11 in story structure. Similar results for bilinguals' two languages were also found by Akinci, Jisa & Kern (2001), in a study of Frog story narratives produced by 42 Turkish-French bilinguals aged 5, 7, and 10 in both languages. In each age group, the children produced equally complex narratives in both Turkish and French, and narratives produced by the older groups were more complex than those of the younger groups. In a study of Frog stories of 12 English-Spanish bilinguals aged 4;0-6;11, Fiestas & Peña (2004), also found that the narratives had similar levels of complexity in both languages.

In sum, results from earlier studies on the development of narrative macrostructure thus indicate the following: (1) There is a development in narrative macrostructure with age within the preschool and early school years; (2) bilinguals perform similarly in their two languages; (3) narrative comprehension develops earlier than production, and (4) certain story grammar components such as attempts and outcomes, which may be visually salient in the stimulus materials, are produced more frequently than other types of story components, such as goals and internal states, which require inferencing. Much is still unknown about (bilingual) children's narrative macrostructure, especially in a Swedish context and especially concerning narrative comprehension. Comparing age groups, comparing the bilinguals' two languages, comparing comprehension and production, and looking closer at different types of components and overall macrostructural complexity, as well as comparing language groups for Swedish and different bilingual groups in their minority languages are all relevant aspects for understanding preschool children's narrative competence. These aspects are therefore analyzed in the current study.

7.1.2 Studies using MAIN

Although MAIN is a relatively new instrument, a number of studies using it to investigate bilingual children's narrative macrostructure have been published in a recent special issue of the journal *Applied Psycholinguistics*.[107] The studies analyzed narratives elicited with MAIN from a range of lan-

[107] In addition to the studies of this special issue, results from Swedish-Russian (Koivistoinen, 2012; Olsson, 2013) and Swedish-French bilinguals (Haessig & Tuvås, 2013) have been reported in unpublished MA theses in speech and language pathology.

guage combinations: Swedish-English (Bohnacker, 2016),[108] Finnish-Swedish (Kunnari, Välimaa, & Laukkanen-Nevala, 2016), German-Russian (Gagarina, 2016), German-Turkish (Maviş, Tunçer, & Gagarina, 2016), Slovak-English (Kapalková, Polišenská, Marková, & Fenton, 2016), and Italian-English (Roch, Florit, & Levorato, 2016).[109] The study by Kunnari et al. (2016) of Finnish-Swedish bilinguals also included a group of monolinguals (speaking the majority language Finnish) for comparison. Notably, no earlier study using MAIN has compared different groups of bilinguals speaking the same majority (or minority) language, as is done in the current study. Whereas all studies included measures of production of macrostructure, only four studies (Bohnacker, 2016; Kapalková et al., 2016; Maviş et al., 2016; Roch et al., 2016) report results for the MAIN comprehension questions (see Section 7.2.1.1 for a description of the questions). Only the study by Bohnacker (2016) explicitly compared comprehension to production. Relatively little is thus known about the relationship between comprehension and production in MAIN. Therefore, the current study compares the children's performance on comprehension and production (Section 7.4).

The focus here is on results from earlier studies using the telling elicitation method which is used in the current study (cf. Section 1.4.1). However, some comments on telling and retelling are necessary. First, it is important to point out that previous studies have not used Cat and Dog for telling, as these stories were originally created to be used for retelling and model story (Gagarina et al., 2012).[110] Thus, all results reported for telling in earlier publications are from Baby Birds/Baby Goats. Second, published results for telling (Baby Birds/Baby Goats) compared to retelling (Cat/Dog) are not consistent across languages and studies. Kunnari et al. (2016) found that the Finnish-Swedish bilinguals had higher scores in retelling than in telling in Finnish, but there was no difference between the two elicitation methods in Swedish. Maviş et al. (2016) found no differences between tell and retell in comprehension or production in Turkish. In these studies, children were tested with both tasks in the same session, with telling always preceding retelling. In Roch et al. (2016), telling and retelling were tested in different sessions, with a counterbalanced order of the tasks. Roch et al. (2016) found that both comprehension and production was better in retelling than in telling, irrespective of language (Italian, English). Third, the studies comparing telling and retelling using the MAIN (Kunnari et al., 2016; Maviş et al.,

[108] This study is based on data originally collected within the frame of three unpublished MA theses in speech and language pathology (Arnoldsson & Aronsson, 2013; Härdelin & Naylor, 2012; Leback & Nilsson, 2012).

[109] Two further studies (Altman, Armon-Lotem, Fichman, & Walters, 2016; Tsimpli, Peristeri, & Andreou, 2016) in the special issue were not included here as they deal with children with SLI and therefore fall outside the scope of the current study.

[110] Model story is an elicitation mode where the child listens to one story (and may also answer probe questions to that story), after which s/he tells another story. When Cat/Dog are used as model story, the child then tells Baby Birds/Baby Goats.

2016; Roch et al., 2016) assumed that Cat/Dog and Baby Birds/Baby Goats are equivalent and that any difference in the children's performance between the two tasks is due to differences in elicitation method. Structurally, Cat/Dog and Baby Birds/Baby Goats are indeed comparable as they all contain the same number of episodes, and they are scored for the same components in comprehension and production (cf. Sections 3.2.2, 7.2.1.1 and 7.3.1.1). It still remains to be shown whether they are equally difficult for the children irrespective of type of elicitation method. If retelling (Cat/Dog) is better than telling (Baby Birds/Baby Goats), it could be the case that children simply perform better on Cat/Dog than on Baby Birds/Baby Goats. In the current study, comprehension and production of macrostructure is compared for telling of both Cat/Dog and Baby Birds/Baby Goats.

All published studies of data from the MAIN used the same composite *story structure score* as laid down in the MAIN scoring protocol. This macrostructure production score (max = 17 points) is also used in the current study. Results on this measure can thus easily be compared between studies. Only two published studies have reported results for different types of *macrostructural components* (Bohnacker, 2016; Kapalková et al., 2016; cf. Section 1.2). Additionally, in some studies, *macrostructural complexity* was investigated. This measure concerns the production of different macrostructural sequences (attempt-outcome, goal-attempt, goal-outcome, goal-attempt-outcome, sometimes 'goal only' is also included here). The exact measures used to analyze complexity vary between studies, making it difficult to compare the results. Some studies (Gagarina, 2016; Roch et al., 2016) assessed the child's best performance, i.e. they assigned one complexity value to each child based on the highest level of complexity reached by the child (i.e. a GAO-sequence counts as more complex than an AO-sequence or a GA/GO-sequence). Other studies classified each episode according to the type of sequence it contained and then compared episodes containing a goal with those that did not (Kunnari et al., 2016). Yet others gave the child a total complexity score based on the production of different types of sequences in the different episodes (Maviş et al., 2016) or only analyzed how often GAO-sequences were produced (Bohnacker, 2016). It should be kept in mind that none of the studies conducted have analyzed macrostructural complexity in exactly the same way as in the current study and that one should therefore be careful when comparing the results.

The above-mentioned six published MAIN-studies (Bohnacker, 2016; Gagarina, 2016; Kapalková et al., 2016; Kunnari et al., 2016; Maviş et al., 2016; Roch et al., 2016) are now described with a focus on age effects and differences between the bilinguals' two languages, after which a general summary of their findings is given.

Bohnacker (2016) analyzed narratives elicited from 52 Swedish-English bilinguals growing up in Sweden, comparing five-year-olds (N=19) with children aged 6–7 (N=33). She analyzed macrostructure production and

comprehension scores as well as different types of components and comprehension questions and complexity, finding no significant differences between the bilinguals' two languages for any of the measures. The 6–7-year-olds performed better than the 5-year-olds in both comprehension and production. There was only an effect of order of testing for the older group, who performed better in the second testing, especially when the second testing was Swedish, which was their stronger language. The children performed better in comprehension than in production, and comprehension was at a very high level already at age 5. Not even the older group produced many goals in their narratives, but all children understood them well when probed. For macrostructural complexity, the five-year-olds produced only 7% GAOs and this increased somewhat by age 6–7 (12%). Only 21% of the narratives of the younger group contained one GAO; the corresponding figure for the 6–7-year-olds was 35%.

Kunnari et al. (2016) studied telling and retelling in 16 Finnish-Swedish bilinguals and 16 Finnish monolinguals aged 5;0–6;7 in a Finnish setting. The bilinguals performed equally well in both languages and there was a clear improvement with (linear) age both for macrostructure production score and macrostructural complexity. For macrostructure production scores in Finnish, there was no difference between the monolinguals and bilinguals in retelling, but in telling the monolinguals had significantly higher scores than the bilinguals. The groups showed comparable macrostructural complexity in Finnish.

In Roch et al. (2016), telling and retelling in both languages of 62 Italian-English bilinguals aged 5–7, growing up in Italy but attending an English-medium school from around age 3, were investigated. For L2 English, there was a large difference between children aged 5–6 (N=30) and those aged 6–7 (N=32), both for comprehension and production. For L1 Italian, there were no differences between the age groups. The younger children performed better in Italian, whereas the older children had similar scores in the two languages.

Gagarina (2016) studied both languages of 58 German-Russian bilinguals growing up in Germany. There were 21 preschoolers (age 2;7–4;4), 15 children in grade 1 (aged 6;5–7;5) and 22 children in grade 3 (aged 7;11–10;6). The preschool children performed significantly lower than the other groups on macrostructure production scores in both languages, but there were no differences between the older groups. In both Russian and German, the preschoolers produced very few GAOs. In German, there was only a difference between the preschoolers and the two older groups in proportions of children producing at least one GAO. In Russian, a higher proportion of the children in Grade 3 produced one or more GAO compared with the children in Grade 1. The children's performances in Russian and German were not compared.

Maviş et al. (2016) conducted two studies of the Turkish of German-Turkish bilinguals, growing up in Germany. In study 1, 36 children aged 3–

7, divided into three age groups (3-year-olds, 4–5-year-olds and 6–7-year-olds), listened to a model story, answered comprehension questions and then told another story with accompanying comprehension questions. Maviş et al. (2016) only found effects of age between the oldest group and the younger ones in macrostructural complexity (as measured by scores on a scale from 0-9 points) and between the youngest and the two older groups in comprehension (both after model story and telling), but no effect on macrostructure production scores. The children performed better on the comprehension questions after model story (Cat/Dog) than after telling (Baby Birds/Baby Goats). Study 2 compared tell and retell for 13 German-Turkish children aged 5;5–7;11. The younger children, aged 5;5–7;0 performed lower than the older group (7;1–7;11) in comprehension, but there were no age effects in production.

Finally, Kapalková et al. (2016) studied comprehension and production in both languages of 40 Slovak L1-English L2 five- to six-year-olds. The children were growing up in Slovakia but attended English-medium preschools and schools. The L2 was always tested in the first session and the L1 in the second testing. The authors chose to combine results for telling and retelling, making it hard to compare their results to other studies. Production and comprehension were compared for the two languages (L1 Slovak, L2 English). For production, scores in the L1 were higher than scores in the L2, but the children scored similarly in comprehension in the two languages. The children produced similar proportions of different components in L1 and L2, with attempts and outcomes being produced much more frequently than the other components. Although in both languages, the proportion of goals in the combined data from tell and retell was relatively high (around 30%), goal comprehension was still much higher (around 70%).

Table 7.1 shows the mean macrostructure production scores for Baby Birds/Baby Goats in the MAIN studies described above (retelling scores for Cat/Dog are not reported). Only scores for age groups within the age range of the current study (i.e. age 4–7) are reported here. Kapalková (2016) only reported combined scores from tell and retell, and as these scores cannot be compared to the scores from the other MAIN studies nor to the current study, they are not included here.

In Table 7.1, we see that there is some variation in scores between studies, although younger groups tend to have lower scores and older groups higher scores. Scores tend to be relatively similar in the two languages (max 1 point difference between the languages, with SDs in the range of 1.5–2.6 points), with the exception of those in the study by Roch et al. (2016), in which the children performed substantially higher in (L1) Italian at age 5–6 (2.8 points difference), but where scores in the two languages were at a similar level at age 6–7. Even the oldest groups of children, 7-year-olds, have mean scores that are not above 50% of the total maximum (17 points).

Table 7.1. Earlier studies using MAIN: Macrostructure production scores (telling), Baby Birds/Baby Goats.

Study	Languages	Age	Mean	SD
Kunnari et al. (2016)	Mono-Finnish	5;0–6;7	6.9	1.7
	Bi-Finnish	5;0–6;7	4.6	1.9
	Bi-Swedish	5;0–6;7	5.4	2.3
Bohnacker (2016)	Bi-Swedish	5	5.7	1.7
		6–7	7.1	2.1
	Bi-English	5	4.9	1.6
		6–7	7.0	2.6
Roch et al. (2016)	Italian L1	5–6	7.1	2.3
		6–7	6.5	2.2
	English L2	5–6	4.3	2.4
		6–7	5.9	2.1
Maviş et al. (2016)	Bi-Turkish	4;0–5;11[a]	5.1	1.8
		6;0–7;11[a]	7.1	2.3
		5;5–7;11[b]	6.2	1.7
Gagarina (2016)	Bi-German	2;7–4;4	3.4	2.2
		6;5–7;5	8.5	1.5
	Bi-Russian	2;7–4;4	3.5	2.5
		6;5–7;5	7.5	2.1

Note. Maximum score = 17 points. All mean scores have been rounded to one decimal.
[a] Narrative told after model story, [b] Narrative told without model.

To summarize the results from earlier studies using the MAIN to assess macrostructure, in line with results for other types of elicitation instruments, there are clear indications that production of macrostructure develops in the preschool period, and that bilingual children score similarly in their two languages. In the current study, we would thus expect a similar outcome, with differences between at least some of the studied age groups, but with similar scores in the bilinguals' two languages. Additionally, the expectation is that the children (age 4–6) will produce relatively few goals and GAO-sequences, but substantial proportions of attempts and outcomes. Based on the results of Bohnacker (2016), we would also expect the children to perform better in comprehension than in production.

7.2 Comprehension of macrostructure

In this section, results on the comprehension of narrative macrostructure are reported. The analyses were carried out on the answers to the comprehension questions of the MAIN for all three languages (Swedish, German, Turkish) for both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats).

In Section 7.2.1, an overview of the methods of scoring and analysis are given. The results are divided into two parts. The first and main part (Section 7.2.2) focuses on the children's overall comprehension *scores* (for MAIN1,

for MAIN2, and for both tasks combined). The second part of the results (Section 7.2.3) concerns the children's *accuracy* on individual comprehension questions. Finally, in Section 7.2.4, the results are summarized. Answers to the following research questions are sought:

- How does comprehension of narrative macrostructure develop with age?
- Are there differences in Swedish narrative comprehension between the language groups?
- Do the bilinguals perform differently in their two languages (Swedish vs German; Swedish vs Turkish)?
- Are there differences between the two minority languages (German and Turkish)?
- Are there differences in comprehension scores between the two narrative tasks MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats)?
- Do the children answer different types of comprehension questions equally well?

7.2.1 Scoring and analysis

7.2.1.1 Scoring of the comprehension questions

Each MAIN story (Cat, Dog, Baby Birds, Baby Goats) is accompanied by ten comprehension questions, targeting the children's understanding of goals, internal states and the general plotline. Each correct answer is awarded 1 point, resulting in a maximum score of 10 points for each story. Since each child answered the questions for two stories (one from MAIN1, one from MAIN2) in each language, the maximum combined macrostructure comprehension score is 20 points per language. The combined score, a measure that is not included in the original MAIN-scoring, was calculated to give a more general view of the child's ability to understand narratives, one that is less dependent on a specific narrative. Questions D1, D4 and D7 probe the child's understanding of the goals of the main story characters in Episode 1, 2, and 3. In Section 7.4, the children's results on these specific questions will be compared with their score on the production of the corresponding goal components for each story (cf. 7.3.1.1). Questions D2 and D5 concern the internal state of the main character for episodes 1 and 2, respectively. Question D8 is a theory of mind question, asking about a hypothetical situation that is not part of the actual story, but that requires understanding of the plot. Questions D3, D6 and D9 are follow-up questions to D2, D5 and D8, respectively; the child can only score a point on the follow-up questions (which are why-questions probing the reason for the character's internal state) if s/he has answered the preceding question correctly. D10, which is a question requiring understanding of the overall plotline, consists of two parts, and the child must answer both correctly to score a point. For D10 on

both MAIN1 and MAIN2, the second part is a why-question. Table 7.2 gives an overview of the ten comprehension questions asked in Cat.[111]

Table 7.2. Overview of the ten comprehension questions, MAIN1 Cat.

Question	Example (*Cat*)
D1. Episode 1 Goal	Why does the cat jump/ leap forward? (picture 1-2)
D2. Episode 1 IST	How does the cat feel? (picture 3)
D3. Episode 1 IST rationale	Why does the cat feel [answer D2]?
D4. Episode 2 Goal	Why does the boy hold the fishing rod in the water? (picture 5)
D5. Episode 2 IST	How does the boy feel? (picture 6)
D6. Episode 2 IST rationale	Why does the boy feel [answer D5]?
D7. Episode 3 Goal	Why does the cat grab the fish? (picture 5)
D8. Theory of Mind IST	Imagine that the boy sees the cat. How does the boy feel? (picture 6)
D9. Theory of Mind IST rationale	Why would the boy feel [answer D8]?
D10. Overall plotline question	Will the boy be friends with the cat? Why?

It is important to point out that what is assessed here is the child's ability to formulate his/her understanding of the story into answers that are both comprehensible and correct. The comprehension questions thus probe the child's ability to verbalize his/her comprehension and thus involve not only comprehension (of the story and of the questions), but also linguistic production. Answering these comprehension questions is thus a task that requires (at least) a basic level of proficiency in the language of testing. Additionally, answering the questions requires certain cognitive abilities, i.e. the actual ability to interpret what is shown in the pictures and draw inferences, e.g. regarding characters' internal states. Thus, although the comprehension questions are less demanding than narrative production, they are not necessarily easy for the child.[112]

The author scored all children's answers to the comprehension questions in Swedish and German. The Turkish answers were scored by Buket Öztekin, a native Turkish-speaking SLP and linguist. There was substantial variation between individual children in the types of answers (i.e. in the different types of formulations used by the children). Since the original MAIN scoring protocol (Gagarina et al., 2012) only covers a limited number of (typical) cases, the need arose for more comprehensive and homogenized scoring across languages. Detailed scoring guidelines (Guidelines for scoring macrostructure in MAIN, version March 2018), based on the MAIN manual (Gagarina et al., 2012) and including authentic examples from English,

[111] For the questions asked in the other MAIN-stories, see Gagarina et al. (2012, Appendix).
[112] This fact has also been repeatedly pointed out by Ute Bohnacker (p.c.).

German, Swedish, and Turkish, were developed within the BiLI-TAS project (Bohnacker, 2013). These scoring guidelines were based on discussions of a large number of unclear cases within the BiLI-TAS project group (including the author of the current study). The project group consisted of native and near-native speakers of the studied languages (Swedish, German, Turkish). After additional discussions in the original team of MAIN authors (which includes the BiLI-TAS PI Ute Bohnacker), the BiLI-TAS guidelines are currently being adapted to function as extended general scoring guidelines for the MAIN. These guidelines were followed closely in the current study. All unclear cases were discussed in meetings with members of the BiLI-TAS team until agreement was reached. Additionally, all Swedish data from the Swedish-Turkish bilinguals was checked for scoring consistency by a native Swedish SLP research assistant (Karin Koltay), who had received extensive training in the scoring procedure. Very few inconsistencies and scoring disagreements were found and these could be solved either by consulting the guidelines or through discussion.

The total number of questions to be asked was 3,320 for Swedish (166 children x 10 questions x 2 stories), 920 for German (46 children x 10 questions x 2 stories), and 960 for Turkish (48 children x 10 questions x 2 stories). In the (relatively few) cases where the experimenter had forgotten to ask the child a comprehension question, following common practice in cases when there is not more than 2% missing data (Widaman, 2006), sample mean substitution was used. This means that the child was given a score for the question for which data was missing that was the same as the mean for that question for the particular age and language group that the child belonged to. To give an example, the monolingual five-year-old MoSwe5-07 was not asked question D7 in MAIN2. This child was given the mean score on D7 for the monolingual five-year-olds as a substitute D7-score. The questions that were most commonly (although still rarely) forgotten by experimenters were D7 and D10 (especially the follow-up question). When data was missing for three or more comprehension questions in one task, the child was not included in the analysis (a total of six cases, see below).

For the monolinguals, there was 0.3% missing data (4/1,440). In the Swedish of the Swedish-German bilinguals, no data was missing. One child (BiGer5-14) answered the MAIN1 questions in German in the Swedish testing and these answers were not analyzed. This child was not included in analyses of Swedish combined comprehension scores or in the comparison of combined scores for the two languages. In the German testings, there was 0.4% missing data (4/910). The results from Swedish for two Swedish-Turkish four-year-olds (BiTur4-06, BiTur4-14) were not included in the analysis of comprehension of narrative macrostructure, since, in both cases, three or more questions had not been asked for both stories. To score the remaining questions (based on sample mean substitution as described above) would give a misleading picture of these children's narrative comprehension.

This means that 46 Swedish-Turkish children out of 48 were included in the analysis. In the Swedish narratives of these 46 Swedish-Turkish bilinguals, there was 1.6% missing data (15/920).[113] In Turkish, MAIN1 data from one five-year-old (BiTur5-21) and MAIN2 data from one four-year-old (BiTur4-26) were excluded from analysis as three or more questions had not been asked. In Turkish, 47 children were thus included in the analysis for MAIN1 and MAIN2, respectively, and 46 children were included in the analyses of the Turkish combined comprehension score. In the comparisons of Swedish and Turkish, 44 Swedish-Turkish bilinguals were included. There was 0.6% missing data (6/940) in Turkish.

7.2.1.2 Statistical analyses

To answer the research questions stated above about differences between age and language groups and between narrative tasks (Section 7.2.2), a number of different analyses were carried out. Combined comprehension scores (i.e. MAIN1 + MAIN2 scores) were correlated with age in months for each language (for Swedish, this analysis was carried out separately for the three language groups). For the bilingual groups, the combined macrostructure comprehension scores (i.e. MAIN1 score + MAIN2 score) in their two languages were correlated, so as to compare their broader narrative comprehension in two languages.

For each language (Swedish, German, Turkish), in order to test effects of the narrative task on performance, the children's scores for MAIN1 and MAIN2 comprehension were compared using paired-samples t-tests. For the bilinguals, their MAIN1 and MAIN2 scores in the two languages were compared using paired-samples t-tests. Additionally, using independent-samples t-tests, scores from the two stories of each narrative task (Cat and Dog for MAIN1; Baby Birds and Baby Goats for MAIN2) were compared. Furthermore, the effects of the order of testing were analyzed for both languages of the bilinguals. Whenever any of these variables (story; order of testing) was significant, it was included in the analyses carried out on the MAIN1 and MAIN2 scores.

To analyze effects of language group and age group on Swedish macrostructure comprehension, two factorial ANOVAs were carried out, one for MAIN1 and one for MAIN2. These analyses included the effect of story in cases where the independent-samples t-tests had shown that there was a significant difference for this variable. For age and language group, post-hoc tests (Bonferroni) were used to find out between which groups there were significant differences.

[113] The higher percentage for the Swedish data from the Swedish-Turkish bilinguals, compared with the monolingual and Swedish-German data was probably due to the size of the team of experimenters; it is reasonable to assume that having a higher number of experimenters could lead to a higher proportion of errors, as each experimenter may make a few errors.

The effects of age group on Turkish and German macrostructure comprehension were analyzed using either factorial or regular ANOVAs with post-hoc tests (Bonferroni), depending on whether the variables story and/or order of testing were significant. The MAIN1 and MAIN2 scores were also correlated with age in months, in order to judge if there was a subtle age effect that could not be picked up by analyzing differences between age groups. Finally, the comprehension scores from Turkish and German were compared using independent-samples t-tests.

For all factorial ANOVAs, whenever a significant interaction was found between two or more independent variables, follow-up analyses were carried out, e.g. in the form of ANOVAs or (independent samples) t-tests, depending on the type of interaction.

For the accuracy percentage on individual comprehension questions (Section 7.2.3), only descriptive statistics are given. As goals are central to the production of well-formed narrative structure (cf. Section 1.1), goal comprehension (in proportion out of the total possible score for goals) was compared for both MAIN1 and MAIN2 for the three groups in Swedish, using the Kruskal-Wallis test (with pairwise comparisons between the three groups). The two bilingual groups' results in the two minority languages were compared using the Mann-Whitney test (German vs Turkish).[114]

7.2.2 Macrostructure comprehension scores

7.2.2.1 Swedish

Table 7.3 shows the results for Swedish macrostructure comprehension for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) as well as the combined comprehension scores from both narratives for the three language groups. In Table 7.4, answers to the MAIN1 comprehension questions are shown, from one low- and one high-performing four-year-old child. Note that the children are of the exact same age; this shows the large variation in performance among the younger children.

[114] As these analyses were carried out on proportions, non-parametric tests were used.

Table 7.3. Swedish macrostructure comprehension scores, MAIN1, MAIN2 and combined, by language group.

	MAIN1	MAIN2	Combined
Monolinguals (N=72)			
Mean (SD)	7.9 (2.1)	6.9 (2.3)	14.7 (3.6)
Range	2 – 10	1 – 10	3 – 20
Swedish-German bilinguals (N=46)			
Mean (SD)	8.4 (1.4)	7.3 (2.4)	15.6 (3.2)
Range	5 – 10	2 – 10	8 – 20
Swedish-Turkish bilinguals (N=48)			
Mean (SD)	7.5 (2.2)	5.5 (2.6)	13.0 (4.4)
Range	1 – 10	0 – 10	2.17 – 19

Note. Max score for MAIN1 and MAIN2 = 10 points. Max score for combined = 20 points.

Table 7.4. Examples of answers to the comprehension questions, MAIN1 Cat.

	MoSwe406, 4;6, Cat		**MoSwe414, 4;6, Cat**	
D1	för den tänkte ta fjärilen 'for it intended to take the butterfly'	1	för att han vill ta fjärilen 'because he wants to take the butterfly'	1
D2	den trillade och så flög fjärilen bort 'it fell and then the butterfly flew away'	0	jätteont 'very painful'	1
D3	-	0	för att han gjorde illa sig 'because he hurt himself'	1
D4	för den skulle dra upp bollen 'for it should pull up the ball'	1	för att skulle fiska upp bollen 'because he should fish up the ball'	1
D5	den tänkte spela 'he intended to play'	0	glad 'happy'	1
D6	-	0	för att han har fiskat upp bollen 'because he has fished up his ball'	1
D7	för katten var hungrig 'because the cat was hungry'	1	för att han blev sugen på fiskarna 'because he got a craving for the fish'	1
D8	vet inte 'don't know'	0	jätteledsen 'very sad'	1
D9	-	0	för att katten äter upp alla 'because the cat is eating all'	1
D10	nej, det var kanske dens katt, kanske den fiskade fiskarna till den katten 'no, it was maybe its cat, maybe it fished the fish for that cat'	0	nej, för att hon åt upp alla fiskarna 'no, because she ate all the fish'	1
	Total points	**3**	**Total points**	**10**

Before analyzing scores for MAIN1 and MAIN2 further, something more should be said about the combined comprehension scores. There were highly significant positive correlations between age and combined comprehension score in all three language groups (monolinguals: $r = .565$, $p < .001$, Swedish-German bilinguals: $r = .649$, $p < .001$; Swedish-Turkish bilinguals: $r = .586$, $p < .001$). These relationships between age and macrostructure comprehension can be seen in Figure 7.1.

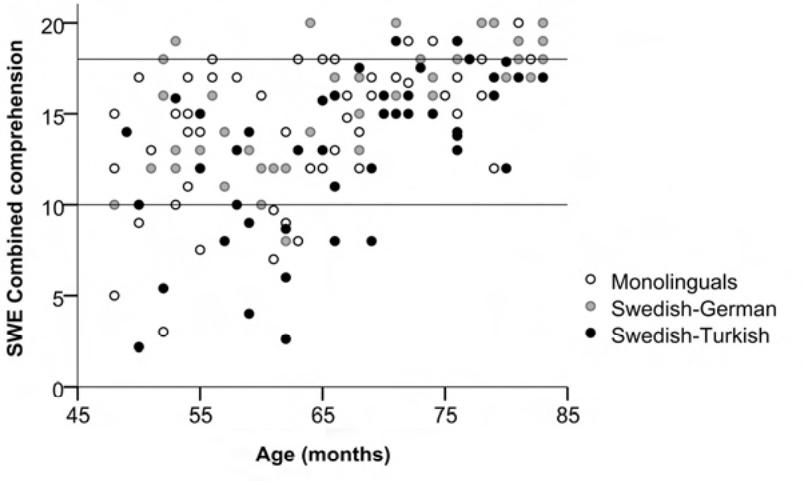

Figure 7.1. Combined (MAIN1 + MAIN2) comprehension of macrostructure, Swedish. Max score = 20 points. A dot may represent more than one individual child. Lines indicate 50% (10 points) and 90% (18 points).

Figure 7.1 also shows that most children scored between 50% (10 points) and 90% (18 points), indicating that most children showed relatively good, but not perfect, comprehension of the two stories. It is also notable that no child above the age of 5;9 scored below 10 points. The lowest score for a six-year-old is 12 points (one monolingual, aged 6;7, one Swedish-Turkish bilingual, aged 6;8); most six-year-olds scored substantially higher than that, especially among the monolinguals and the Swedish-German bilinguals. Only eight out of 163 children, two monolinguals and six Swedish-German bilinguals, scored at maximum (20 points). Thus, for most children, comprehension as measured by the MAIN comprehension questions is good but not fully developed by age 6.

Figure 7.2 shows the comprehension scores for MAIN1 and MAIN2 separately by language and age group.

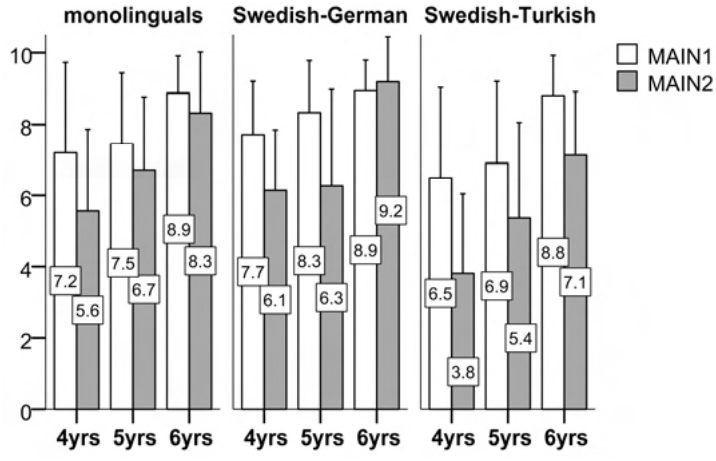

Figure 7.2. Mean macrostructure comprehension scores, Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), by language and age group. Max score = 10 points. Error bars show +1SD.

There was a significant difference in mean scores between the tasks ($t(162) = 6.906$, $p < .001$); overall, the children scored better on MAIN1 (M = 7.9, SD = 2.0) than on MAIN2 (M = 6.6, SD = 2.5). As can be seen in Figure 7.2, the difference for the monolingual six-year-olds was small, and the Swedish-German six-year-olds even scored slightly higher on MAIN2 (M = 9.2, SD = 1.3) than on MAIN1 (M = 8.9, SD = 0.9). On the other hand, the scores of the Swedish-Turkish four-year-olds on MAIN1 (M = 6.5, SD = 2.5) were substantially higher than on MAIN2 (M = 3.8, SD = 2.2). In fact, the Swedish-Turkish four-year-olds showed low comprehension in MAIN2.[115] Figure 7.2 also illustrates that the variation was smaller in the older groups; this is especially evident when comparing the standard deviations of the four- and six-year-olds in all three language groups.

Overall, there was no significant difference in mean scores between Cat (M = 7.9, SD = 2.0) and Dog (M = 7.8, SD = 1.9) ($t(161) = .256$, $p = .792$). Mean scores were significantly lower for Baby Birds (M = 6.1, SD = 2.3) than for Baby Goats (M = 7.1, SD = 2.5) ($t(162) = -2.863$, $p = .005$). This means that the variable story needs to be included in the statistical analysis for MAIN2, but not for MAIN1.

To test the effect of the two independent variables language group and age group on the MAIN1 comprehension scores, an Age group x Language group (3x3) factorial ANOVA was carried out. There was a significant main effect of age group ($F(2, 154) = 12.822$, $p < .001$, $\eta_p^2 = .143$). The post-hoc

[115] This does not necessarily mean that they failed to comprehend the story content, only that they were not able to verbalize it in a comprehensible manner, i.e. answer the comprehension questions correctly. This could be due to limited vocabulary and/or grammar.

tests showed that there was a significant difference between the six-year-olds and both the two younger groups, but no significant difference between the four- and five-year-olds.[116] The main effect of language group was not significant ($F(2, 154) = 2.888$, $p = .059$, $\eta_p^2 = .036$). There was no significant interaction effect ($F(4, 154) = .599$, $p = .664$, $\eta_p^2 = .015$), which indicates the effects of age were the same for each language group, and that there was no effect of language group in any age group. This confirms the picture shown in Figure 7.2.

For MAIN2, an Age group x Language group x Story (3x3x2) factorial ANOVA was run. There were significant main effects of age group ($F(2, 146) = 33.516$, $p < .001$, $\eta_p^2 = .315$), and of language group ($F(2, 146) = 11.037$, $p < .001$, $\eta_p^2 = .131$). The post-hoc tests for age group showed significant differences between all three age groups; the six-year-olds performed better than the five-year-olds, who in their turn performed better than the four-year-olds. Additionally, the Swedish-Turkish children scored significantly lower than the other two language groups. There were no significant differences between the Swedish-German bilinguals and the monolinguals.[117]

Additionally, there was a main effect of story ($F(1, 146) = 11.320$, $p = .001$, $\eta_p^2 = .072$); scores on Baby Goats were significantly higher than those on Baby Birds. There was also a significant two-way interaction between language group and story ($F(2, 146) = 3.105$, $p = .048$, $\eta_p^2 = .041$). This interaction, indicating that the effect of story was not the same in all language groups, is shown in Figure 7.3.

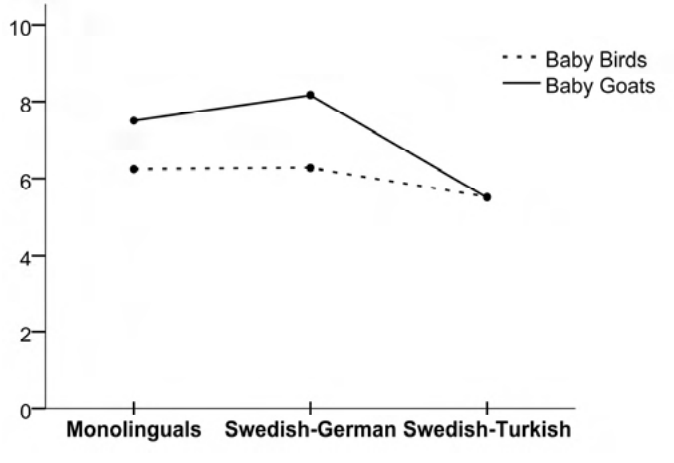

Figure 7.3. Swedish MAIN2 comprehension scores, two-way interaction Language group x Story. Max score = 10 points.

[116] For p-values of the pairwise comparisons between the age groups in MAIN1, see Appendix 5, Table A5.2.

[117] P-values of all pairwise comparisons between age groups and language groups for MAIN2, can be found in Appendix 5, Table A5.1 and Table A5.2, respectively.

Figure 7.3 clearly shows that whereas scores for Baby Goats are higher than for Baby Birds in the monolingual group and the Swedish-German group, the Swedish-Turkish bilinguals scored identically on both stories. Independent-samples t-tests run on the language groups separately confirm this interpretation: there was a significant difference between Baby Birds and Baby Goats for the monolinguals ($t(70) = -2.398$, $p = .019$), and for the Swedish-German bilinguals ($t(44) = -2.898$, $p = .006$), but not for the Swedish-Turkish bilinguals ($t(44) = .030$, $p = .976$). Additionally, scores for the three language groups on Baby Birds are far more similar than for Baby Goats. Running one-way ANOVAs with language group as the independent variable on the two stories separately revealed a significant difference between the language groups for Baby Goats ($F(2, 80) = 8.383$, $p < .001$), but not for Baby Birds ($F(2, 78) = .806$, $p = .450$). Post-hoc tests showed that the only significant difference for Baby Goats was between the Swedish-Turkish bilinguals and the other two groups. This means that the main effect of language group shown above stems from differences between the language groups on Baby Goats.[118]

In addition to the two-way interaction described above, the three-way interaction Language group x Age group x Story was also significant ($F(4, 146) = 2.763$, $p = .03$, $\eta_p^2 = .070$). This interaction indicates that the effect of age for the two stories is not the same in all the language groups, as can be seen in Figure 7.4.

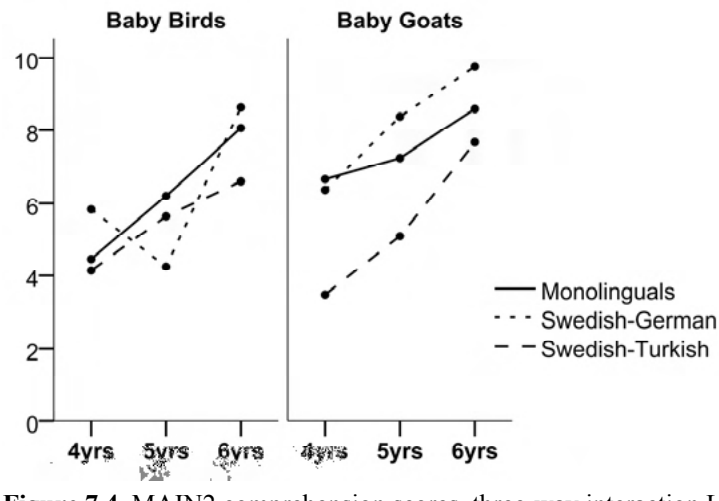

Figure 7.4. MAIN2 comprehension scores, three-way interaction Language group x Age group x Story. Max score = 10 points.

To follow up on the results seen in Figure 7.4, i.e. to be able to analyze the relationship between age group and story further, factorial ANOVAs (with

[118] For p-values of all pairwise comparisons, see Appendix 5, Table A5.1.

age group and story as predictors) were run for the three language groups separately. These analyses revealed that, as expected, there was only a significant interaction between age and story for the Swedish-German bilinguals, due to the different performance of the five-year-olds on Baby Birds and Baby Goats.[119] Since the performance of the Swedish-German five-year-olds on Baby Birds was surprisingly low, a closer look was taken at the scores of these children.[120] All the lowest scores from the Swedish-German five-year-olds came from Baby Birds. It could be the case that, as an accidental result of the sampling, the lowest-performing children in this group all received Baby Birds or that there is an interaction between the child and the story, causing these children to perform lower than they otherwise would have. If their unusually low performance were caused by properties of the story itself, this would not explain why the same radical effect of story is not present in any other group.

Next, the children's performance on the different questions for Baby Goats was investigated more in detail, to see if it was possible to find an explanation for the lower performance of the Swedish-Turkish bilinguals on this story. It could be the case that Swedish-Turkish bilinguals had larger difficulties with some specific questions than with others. The Swedish-Turkish bilinguals scored consistently lower on all comprehension questions of Baby Goats. However, the difference was much larger for some of the questions. Those were mainly the questions from Episode 1, e.g. D1 (22% correct compared to 67% for the monolinguals). Thus, understanding Episode 1 of Baby Goats was especially difficult for the Swedish-Turkish bilinguals. There was also a substantial difference between the groups for question D10. This question requires understanding of the whole plotline and of the characters' roles in the plot. Only around half of the Swedish-Turkish children (48%) answered this question correctly, compared with 86% and 79% of the monolinguals and Swedish-German bilinguals, respectively.

To summarize the results of the Swedish macrostructure comprehension, the children scored significantly better on MAIN1 (Cat/Dog) than on MAIN2 (Baby Birds/Baby Goats). The three language groups scored similarly on MAIN1 (Cat/Dog), and for this narrative task, there was only a significant difference between the six-year-olds and the younger groups. There was no difference between Cat and Dog. For MAIN2, there was a clear age effect, with significant differences between all three age groups. Additionally, the Swedish-Turkish children scored lower than the other two language

[119] Monolinguals: age: $F(2, 66) = 12.332$, $p < .001$, story: $F(1, 66) = 7.643$, $p = .007$, age x story: $F(2, 66) = 1.193$, $p = .310$; Swedish-German bilinguals: Age: $F(2, 40) = 20.147$, $p < .001$, story: $F(1, 40) = 18.531$, $p < .001$, age x story: $F(2, 40) = 6.214$, $p = .004$; Swedish-Turkish bilinguals: age: $F(2, 40) = 7.823$, $p = .001$, story: $F(1, 40) = .005$, $p = .942$, age x story: $F(2, 40) = .686$, $p = .510$.

[120] Note that there were only eight children in this subgroup.

groups. However, this difference was only seen for the Baby Goats story; for Baby Birds, there was no difference between the language groups.

7.2.2.2 Swedish-German bilinguals

In Table 7.5, the Swedish-German bilinguals' results for macrostructure comprehension in Swedish and German are shown.

Table 7.5. Macrostructure comprehension scores, Swedish-German bilinguals (N=46). MAIN1, MAIN2 and combined, by language.

	MAIN1	MAIN2	Combined
Swedish			
Mean (SD)	8.4 (1.4)	7.3 (2.4)	15.6 (3.2)
Range	5 – 10	2 – 10	8 – 20
German			
Mean (SD)	7.9 (2.0)	6.9 (2.3)	14.8 (3.8)
Range	2 – 10	2 – 10	6 – 20

Note. Max score for MAIN1 and MAIN2 = 10 points. Max score for combined = 20 points.

Although mean scores were somewhat higher in Swedish, there were no significant differences between the languages (MAIN1: $t(44) = 1.634$, $p = .109$); MAIN2: $t(45) = .846$, $p = .402$). This result points towards narrative comprehension being a general ability that is less dependent on e.g. lexicon or grammar in a specific language (cf. Bohnacker, 2016; Pearson, 2002). Supporting this interpretation, scores for MAIN1 and MAIN2 in the two languages were significantly correlated (MAIN1: $r = .349$, $p = .019$; MAIN2: $r = .393$, $p = .007$); for both narrative tasks, those children who performed higher on that task in one language also did so in the other language.

There was also a significant positive correlation between combined scores in the two languages ($r = .409$, $p = .005$), as can be seen in Figure 7.5. Figure 7.5 also shows that most Swedish-German children scored above 50% in both languages. Only one child (BiGer5-04) scored below 50% in both languages. Three children scored below 50% in German and also relatively low in Swedish. Five further children scored relatively low in German (i.e. around 50%), but at least 75% in Swedish.

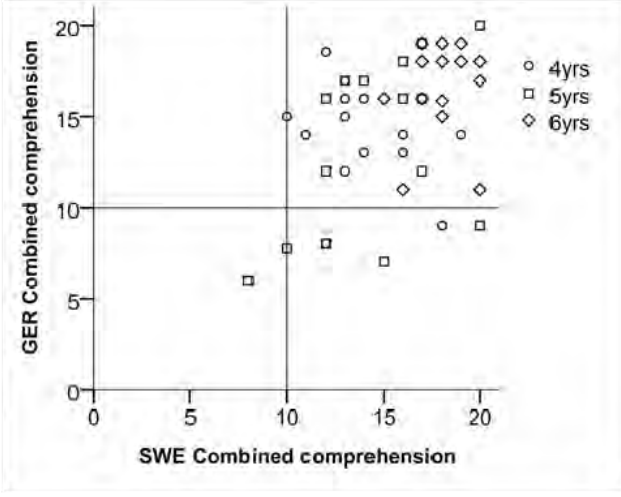

Figure 7.5. Combined (MAIN1 + MAIN2) macrostructure comprehension scores, Swedish and German. Max score = 20 points. A dot may represent more than one individual child. Lines on both axes indicate 50% (= 10 points).

There was a significant positive correlation between age in months and combined scores in German (r = .429, p = .003). This correlation can be seen clearly in Figure 7.6.

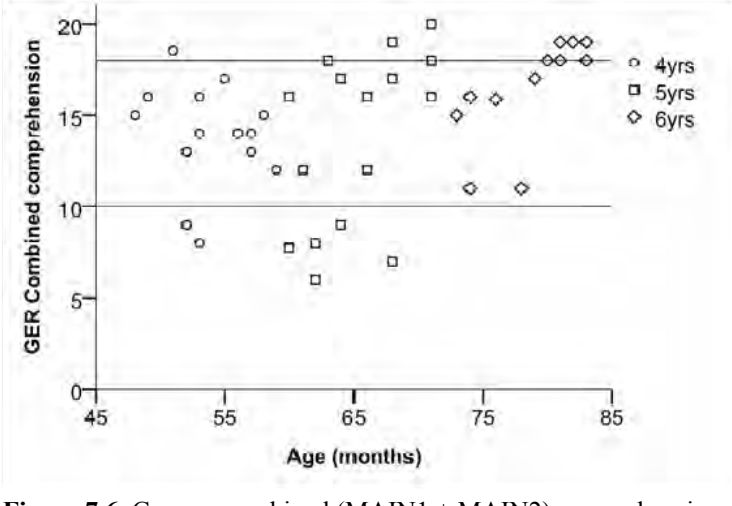

Figure 7.6. German combined (MAIN1 + MAIN2) comprehension scores and age (months). Max score = 20 points. A dot may represent more than one individual child. Lines indicate 50% (= 10 points), and 90% (= 18 points).

The lowest-scoring children were all four- and five-year-olds. All six-year-olds except two scored 15 points or higher. Just as was the case for Swedish (cf. Figure 7.1), most children scored between 50% and 90%. Only one child

out of 46 scored full points on the German comprehension, indicating that comprehension of macrostructure in MAIN is not fully developed by age 6.

Figure 7.7 shows the mean MAIN1 and MAIN2 comprehension scores for Swedish and German by age group.

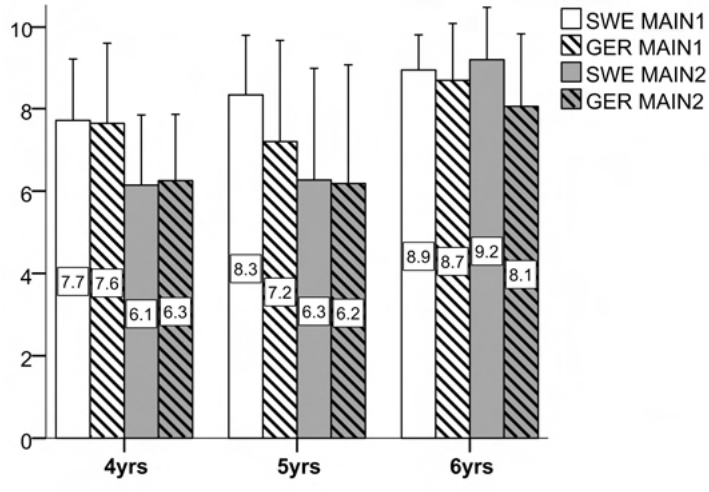

Figure 7.7. Mean macrostructure comprehension scores, German and Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), Swedish-German bilinguals, by age group. Max score = 10 points. Error bars show +1 SD.

In German, just as in Swedish, there was a significant difference between MAIN1 and MAIN2 ($t(45) = 2.972$, $p = .005$); the children generally scored higher on MAIN1 than on MAIN2 (cf. Table 7.4). As can be seen in Figure 7.7, the difference between the narrative tasks was smaller for the six-year-olds than for the younger groups. These results are very similar to the results for Swedish presented above (Section 7.2.2).

The independent-samples t-tests showed that the Swedish-German children performed equally well on Cat (M = 7.8, SD = 1.9) and on Dog (M = 7.9, SD = 2.2) of the German MAIN1 ($t(44) = -.145$, $p = .885$). Although there was a relatively large difference in mean scores between Baby Birds (M = 7.4, SD = 2.2) and Baby Goats (M = 6.4, SD = 2.4), this difference was not significant ($t(44) = 1.406$, $p = .167$).

For Swedish, there was no effect of the order of testing. On both narrative tasks, the children performed equally well irrespectively of whether Swedish was tested first or not (MAIN1: $t(43) = 1.529$, $p = .134$; MAIN2: $t(44) = .244$, $p = .808$). As there is no order effect and analyses of the effect of age have already been carried out on the Swedish data from all three language groups, no further analyses are carried out here on the Swedish data from the Swedish-German bilinguals.

For German on the other hand, there was a clear order effect for both MAIN1 ($t(44) = 3.233$, $p = .002$) and MAIN2 ($t(44) = 2.658$, $p = .011$). The

children performed significantly better when German was the language of the second testing (MAIN1: M = 8.7, SD = 1.4; MAIN2: M = 7.8, SD = 1.9) than when it was tested first (MAIN1: M = 7.0, SD = 2.2; MAIN2: M = 6.1, SD = 2.4). The difference between scores for German in the first or in the second testing was relatively large for both narrative tasks. Analyses of age effects on the comprehension scores thus need to take order of testing into account, for both narrative tasks.

For the German MAIN1 macrostructure comprehension, the Age group x Order factorial (3x2) ANOVA showed a clear main effect of order of testing ($F(1, 40) = 12.117$, $p = .001$, $\eta_p^2 = .232$), but the main effect of age just failed to reach significance ($F(2, 40) = 3.054$, $p = .058$, $\eta_p^2 = .132$). There was no significant interaction effect between age group and order of testing ($F(2, 40) = 2.409$, $p = .103$, $\eta_p^2 = .107$). Although the factorial ANOVA did not show any effect of age group, there was a significant correlation between MAIN1 comprehension scores and age in months ($r = .333$, $p = .024$).

For the German MAIN2 macrostructure comprehension, the same type of factorial ANOVA was run. The resulting model showed significant main effects of both order of testing ($F(1, 40) = 7.083$, $p = .008$, partial $\eta^2 = .163$) and age group ($F(2, 40) = 3.581$, $p = .037$, $\eta_p^2 = .152$), but no significant interaction effect ($F(2, 40) = .543$, $p = .585$, $\eta_p^2 = .026$). Although the main effect of age group was significant, the post-hoc tests did not show any significant differences between individual age groups.[121] However, there was a significant correlation with age in months ($r = .411$, $p = .005$).

To summarize, there were no differences between German and Swedish in comprehension. In German, the children performed better on MAIN1 (Cat/Dog) than MAIN2 (Baby Birds/Baby Goats), and there was also a clear effect of the order of testing for German, with children having the second testing in German performing better on both narrative tasks. There were clearer effects of age on MAIN2 than on MAIN1, and when age in months was analyzed and not age groups. For these specific measures, MAIN1 and MAIN2 comprehension scores, and for the minority language German of the Swedish-German bilinguals, it thus seems to be the case that a comparison of the three age groups does not capture the development with linear age that is present in the data.

7.2.2.3 Swedish-Turkish bilinguals

In Table 7.6, results for the Swedish-Turkish bilinguals' macrostructure comprehension in Swedish and Turkish are shown.

[121] P-values for the pairwise comparison between the age groups can be found in Appendix 5, Table A5.2.

Table 7.6. Macrostructure comprehension scores, Swedish-Turkish bilinguals (N=48), MAIN1, MAIN2 and combined, by language.

	MAIN1	MAIN2	Combined
Swedish			
Mean (SD)	7.5 (2.2)	5.5 (2.6)	12.9 (4.3)
Range	1 – 10	0 – 10	2.17 – 19
Turkish			
Mean (SD)	6.9 (2.4)	5.8 (2.8)	12.9 (4.8)
Range	1 – 10	0 – 10	1 – 20

Note. Max score for MAIN1 and MAIN2 = 10 points. Max score for combined = 20 points.

In both languages, the Swedish-Turkish bilinguals' score range is wide, with children scoring both at floor and at ceiling. Surprisingly, there were no significant differences between the languages (MAIN1: $t(44) = 1.275$, $p = .209$; MAIN2: $t(44) = -.572$, $p = .570$). Although the Swedish-Turkish children were more proficient in Turkish in terms of e.g. vocabulary production (cf. Section 5.2.4), they did not score higher on the Turkish comprehension of macrostructure. This may indicate that narrative comprehension is a general cognitive ability that is less dependent on language proficiency, a similar conclusion to the one drawn from the results for the Swedish-German bilinguals. Scores on MAIN1 in Swedish and Turkish were not significantly correlated ($r = .160$, $p = .295$), but Turkish MAIN2 scores correlated significantly with Swedish MAIN2 ($r = .388$, $p = .009$). Combined scores in the two languages were also significantly correlated ($r = .329$, $p = .029$), as can be seen in Figure 7.8.

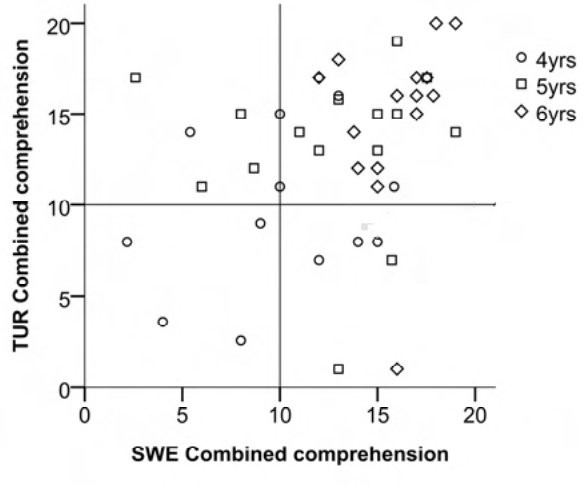

Figure 7.8. Combined (MAIN1 + MAIN2) macrostructure comprehension scores, Swedish and Turkish. Max score = 20 points. A dot may represent more than one individual child. Lines on both axes indicate 50% (= 10 points).

The majority of the Swedish-Turkish children scored above 50% in both languages. Only four four-year-olds scored below 50% in both languages. Further, a number of children had a substantially higher performance in one language: five children scored below 50% in Swedish, but not in Turkish and six children scored below 50% in Turkish, but not in Swedish. With the exception of one six-year-old who was not able to answer more than one question correctly in Turkish, all the low-scoring children were four- and five-year-olds.

There was a significant correlation between age in months and combined scores in Turkish (r = .392, p = .007). This correlation can be seen clearly in Figure 7.9.

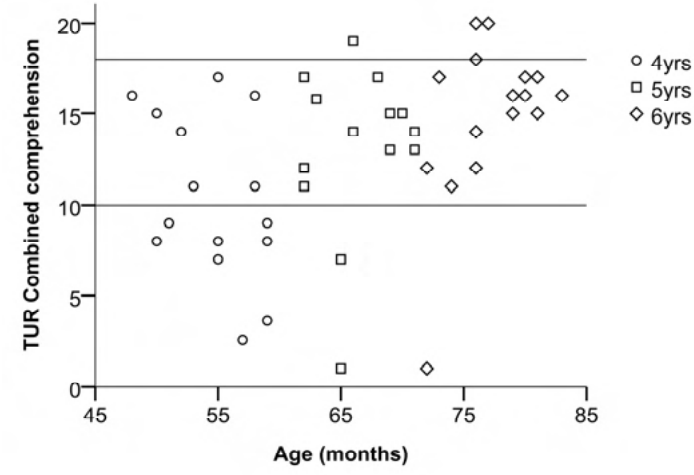

Figure 7.9. Turkish combined (MAIN1 + MAIN2) comprehension scores and age (months). Max score = 20 points. A dot may represent more than one individual child. Lines indicate 50% (= 10 points), and 90% (= 18 points).

Figure 7.10 shows the mean MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) scores for Swedish and Turkish by age group. In Turkish, just as in Swedish and German, there was a significant difference between MAIN1 and MAIN2 ($t(45) = 3.451$, p = .001); the children generally scored higher on MAIN1 comprehension than on MAIN2 (cf. also Table 7.5). There was also a significant correlation between scores on Turkish MAIN1 and MAIN2 (r = .690, $p < .001$); children who scored high on one story, also did so on the other. Differences between MAIN1 and MAIN2 were large for the four-year-olds, who performed equally low on MAIN2 in both languages, but smaller for the older groups.

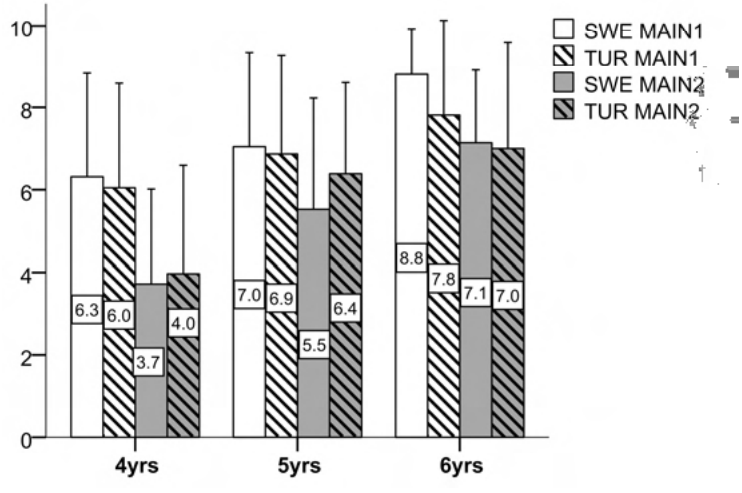

Figure 7.10. Mean macrostructure comprehension scores, Turkish and Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), Swedish-Turkish bilinguals, by age group. Max score = 10 points. Error bars show +1 SD.

In Turkish, mean scores for Cat (M = 6.6, SD = 2.1) and Dog (M = 7.3, SD = 2.8) were not significantly different ($t(45) = -.978$, $p = .334$). Although scores for Baby Goats (M = 6.4, SD = 2.3) were substantially higher than for Baby Birds (M = 5.3, SD = 3.2), these differences were not large enough to reach significance ($t(45) = -1.381$, $p = .175$).

There was no effect of order of testing on the Swedish-Turkish children's comprehension of macrostructure, neither for Turkish (MAIN1: $t(45) = 1.108$, $p = .274$; MAIN2: $t(45) = .953$, $p = .346$), nor for Swedish (MAIN1: $t(44) = 1.050$, $p = .299$; MAIN2: $t(44) = .988$, $p = .329$); the Swedish-Turkish bilinguals performed equally well in both languages irrespective of which language was tested first.

Since there were no significant effects of story or of order of testing, one-way ANOVAs were used to determine if there were any differences between the age groups on the Turkish MAIN1 and MAIN2 comprehension scores. There were no significant differences between the age groups for MAIN1 ($F(2, 44) = 2.297$, $p = .112$), but for MAIN2 there were ($F(2, 44) = 4.326$, $p = .019$). Post-hoc tests showed that the four-year-olds performed significantly lower than the six-year-olds on MAIN2, but that the five-year-olds did not perform differently from the two other groups.[122] There were significant correlations with age in months, both for MAIN1 ($r = .333$, $p = .022$) and for MAIN2 ($r = .386$, $p = .007$). For Turkish, just as for Swedish and German, we thus see clearer effects of age in narrative comprehension for MAIN2 than for MAIN1.

[122] The p-values for all pairwise comparisons can be found in Appendix 5, Table A5.2.

7.2.2.4 Comparing German and Turkish

In Table 7.7, scores for the Swedish-German children on the German comprehension and for the Swedish-Turkish bilinguals on the Turkish comprehension are shown.

Table 7.7. Macrostructure comprehension scores, German and Turkish.

	MAIN1	MAIN2
Swedish-German bilinguals (German, N=46)		
Mean (SD)	7.9 (2.0)	6.9 (2.3)
Range	2 – 10	2 – 10
Swedish-Turkish bilinguals (Turkish, N=48)		
Mean (SD)	6.9 (2.4)	5.8 (2.8)
Range	1 – 10	0 – 10

Note. Max score = 10 points.

There was a clear significant difference in scores between the German of the Swedish-German bilinguals and the Turkish of the Swedish-Turkish bilinguals; on both MAIN1 ($t(88.748) = 2.097$, $p = .039$) and MAIN2 ($t(88.229) = 2.148$, $p = .034$), the Swedish-German bilinguals performed better in German than the Swedish-Turkish bilinguals did in Turkish. The variation was larger in the Turkish results, as can be seen by the higher SDs for both MAIN1 and MAIN2, though ranges are wide in both groups.

The Swedish-German children thus performed better than the Swedish-Turkish children in both Swedish (see Section 7.2.2.1) and the minority language. For Swedish, this is not surprising, as the Swedish-German children are more proficient in Swedish than the Swedish-Turkish children with higher CLT scores (cf. Section 5.2.2) and higher proficiency ratings (cf. Section 3.1.2.3). The overall impression is that the Swedish-Turkish children are more proficient in Turkish than the Swedish-German children are in German, especially if one looks at grammatical correctness.[123] It is thus surprising that the Swedish-Turkish children performed lower on narrative comprehension in Turkish than the Swedish-German children did in German. There may be different explanations for the Swedish-German children's better performance, such as similarities between Swedish and German and the ability to adapt to the experimental setting. We will return to these explanations in the discussion below (Section 7.5).

[123] Note that, in both language groups, a smaller group of children have relatively limited proficiency in the minority language.

7.2.3 Individual comprehension questions

Table 7.8 shows the percentage correct answers on the individual comprehension questions of the Swedish, German and Turkish MAIN1 (Cat/Dog). The pattern in Swedish is the same in all three language groups, with slight variations between the groups in the scores on individual questions.[124] The children performed relatively well on all Swedish MAIN1 comprehension questions except D10. Accuracy on the other nine questions ranges from 72% to 94%, but only 33% of the children scored a point on D10. Even the Theory of mind question D8 and its follow-up question D9 were easy for most of the children. The pattern found in German is identical to the Swedish one, with similarly low accuracy on D10 (37%) and high accuracy (74%–98%) on the other questions. For Turkish, scores are generally somewhat lower than for the other two languages, especially on the D8+D9 questions.

Table 7.8. MAIN1 (Cat/Dog) comprehension, percentage (%) correct answers on the ten comprehension questions, Swedish, German and Turkish

Question	Swedish (N = 163)	German (N = 46)	Turkish (N = 48)
D1. Episode 1 Goal	85	80	79
D2. Episode 1 IST	79	83	81
D3. Episode 1 IST rationale	72	74	70
D4. Episode 2 Goal	86	74	67
D5. Episode 2 IST	93	93	89
D6. Episode 2 IST rationale	86	93	81
D7. Episode 3 Goal	94	98	83
D8. Theory of Mind IST	82	80	60
D9. Theory of Mind IST rationale	77	74	53
D10. Overall plotline question	33	37	28

In Table 7.9, results from the Swedish, German and Turkish MAIN2 (Baby Birds/Baby Goats) questions are shown. The patterns found in Swedish and German are similar, with relatively low scores on the Theory of mind questions (D8+D9) as well as on the follow-up question about IST in Episode 1 (D3). The main difference between the language groups in Swedish was that the Swedish-Turkish bilinguals scored consistently lower than the other two groups on all questions (see Appendix 5, Table A5.4). Other differences between the three language groups on specific Swedish MAIN2 questions have been described above (see Section 7.2.2.1). In Turkish, the children experienced difficulties with the same questions as in Swedish and German, but additionally, accuracy percentages were low on the question about the

[124] The results per language group, showing the same pattern, can be found in in Appendix 5, Table A5.3.

goal of episode 3 (D7) as well as the question concerning the overall plotline (D10).

Table 7.9. MAIN2 (Baby Birds/Baby Goats) comprehension, percentage (%) correct answers on the ten comprehension questions, Swedish, German and Turkish.

Question	Swedish (N = 164)	German (N = 46)	Turkish (N = 48)
D1. Episode 1 Goal	71	74	72
D2. Episode 1 IST	71	78	70
D3. Episode 1 IST rationale	53	57	47
D4. Episode 2 Goal	88	98	81
D5. Episode 2 IST	90	87	83
D6. Episode 2 IST rationale	76	72	66
D7. Episode 3 Goal	65	80	34
D8. Theory of Mind IST	46	44	53
D9. Theory of Mind IST rationale	32	33	26
D10. Overall plotline question	69	72	47

Goals are one of the most central aspects of more complex narrative structures (cf. Section 1.1). Figure 7.11 shows the results for comprehension of goals (in percentages)[125] for Swedish MAIN1 and MAIN2, by age and language group.

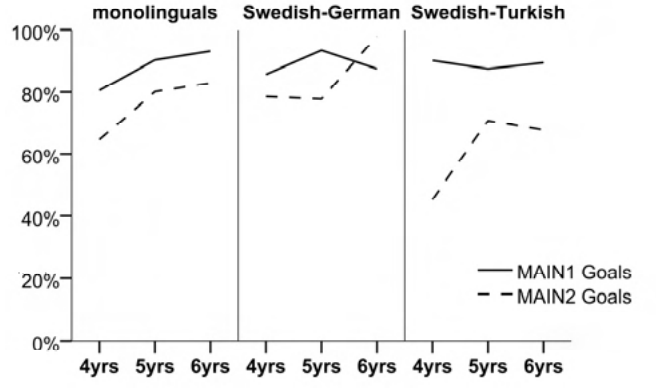

Figure 7.11. Swedish goal comprehension (% correct answers), MAIN1 and MAIN2, by age and language group.

The comprehension of goals in MAIN1 (Cat/Dog) is very high in all three language groups already at age 4, with some increase with age for the monolinguals, and similar scores for the different bilingual age groups. For MAIN2 (Baby Birds/Baby Goats), the comprehension of goals appears to be lower in the Swedish-Turkish group than in in the other two groups. The

[125] This is thus the combined accuracy for the three goal questions D1, D4 and D7.

difference between MAIN1 and MAIN2 goals is large for the Swedish-Turkish four-year-olds, with comprehension of MAIN1 goals at 90% and only 45% of the goals comprehended correctly in MAIN2. Similarly to the results for the overall scores (Section 7.2.2.1), the Kruskal-Wallis test showed no significant differences between the language groups in MAIN1 goal comprehension ($H(2) = 0.012$, $p = .994$). For MAIN2 goal comprehension, there was an effect of language group ($H(2) = 15.186$, $p = .001$). Pairwise comparisons revealed that the Swedish-German bilinguals performed better than the Swedish-Turkish bilinguals, but that there was no difference between the monolinguals and the other two groups. This means that although the Swedish-Turkish bilinguals scored significantly lower than the monolinguals on the MAIN2 overall comprehension score, there was no difference between these groups for goal comprehension.[126]

Age patterns for MAIN2 in the three language groups are different, with most of the increase in percentage correctly comprehended goals found between age 4 and 5 for the monolinguals and the Swedish-Turkish bilinguals, whereas the Swedish-German four- and five-year-olds showed identical performance. The low performance of the Swedish-Turkish bilinguals on MAIN2 goal comprehension is not surprising, considering their overall lower performance on macrostructure comprehension for this narrative task. However, this does not explain the lack of increase in accuracy between age 5 and 6.

Figure 7.12 shows the results for comprehension of goals (in percentages) for the German and Turkish MAIN1 and MAIN2, by age group.

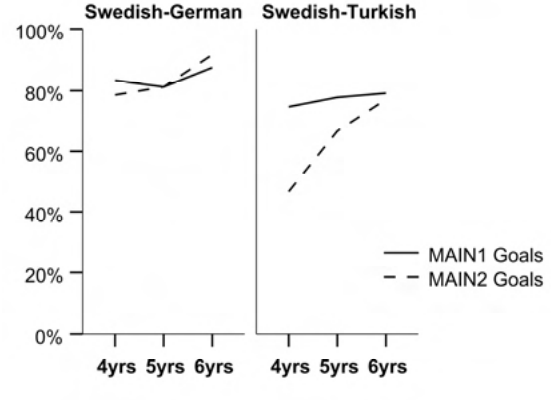

Figure 7.12. Minority language goal comprehension (% correct answers), MAIN1 and MAIN2 by age group, Swedish-German bilinguals (German) and Swedish-Turkish bilinguals (Turkish).

[126] P-values of the pairwise comparisons can be found in Appendix 5, Table A5.5.

For German, the Swedish-German bilinguals showed very good understanding of goals in all age groups and for both narratives. For Turkish, the Swedish-Turkish bilinguals showed relatively good understanding with no improvement with age in MAIN1 (Cat/Dog), results that were similar, although slightly lower, to their Swedish results. Results for Turkish MAIN2 (Baby Birds/Baby Goats) goal comprehension instead show a steep increase, both between age 4 and 5, and between age 5 and 6. Just as in Swedish, the Swedish-Turkish four-year-olds showed a low performance on MAIN2 goal comprehension (47%), compared with MAIN1 (72%). The six-year-olds showed similar results for the two Turkish narratives.

Comparing the two bilingual groups' goal comprehension in the minority language revealed that the Swedish-German bilinguals performed significantly better than the Swedish-Turkish bilinguals in MAIN2 ($U = 659.5$, $p < .001$), but not in MAIN1 ($U = 915.5$, $p = .155$). Thus, even though the Swedish-German bilinguals had a higher overall score on the MAIN1 comprehension questions in German than the Swedish-Turkish bilinguals did in Turkish, both groups comprehended *goals* equally well.

7.2.4 Narrative comprehension: Summary

In all three languages, Swedish, German and Turkish, there were clearer effects of age on MAIN2 (Baby Birds/Baby Goats) than on MAIN1 (Cat/Dog). In Swedish, the six-year-olds scored better than the younger groups on MAIN1, whereas there were differences between all age groups on MAIN2. In German and Turkish, there were only differences between the age groups on MAIN2, although scores on MAIN1 were also significantly correlated with age in months. There were differences between the two tasks. The children scored significantly better on MAIN1 (Cat/Dog) than on MAIN2 (Baby Birds/Baby Goats). In Swedish, the three language groups, monolinguals, Swedish-German bilinguals and Swedish-Turkish bilinguals, scored equally well on MAIN1, but the Swedish-Turkish bilinguals performed lower than the other two language groups on MAIN2. A closer investigation revealed that there was only a significant difference between the language groups on Baby Goats, but not on Baby Birds. The Swedish-Turkish bilinguals performed similarly on both MAIN2 narratives, whereas the children in the other language groups performed better on Baby Goats. For MAIN1, there was no difference between scores on Cat and Dog.

The bilinguals performed similarly in both their languages. In German, there was an effect of the order of testing, with children having German in the second testing performing better on both narrative tasks. No such effect was found for Turkish. On both narrative comprehension tasks, the Swedish-German bilinguals performed better in German than the Swedish-Turkish bilinguals did in Turkish.

The second part of the analysis focused on the accuracy of responses to different comprehension questions, and especially on comprehension of goals. It was found that overall patterns for the different questions were similar for the three languages, with a few exceptions. Certain questions were harder for the children, as shown by lower response accuracy.

There were no significant differences between the language groups in Swedish MAIN1 (Cat/Dog) goal comprehension. In Swedish MAIN2 (Baby Birds/Baby Goats) goal comprehension, the Swedish-German bilinguals performed better than the Swedish-Turkish bilinguals, but there was no difference between the monolinguals and the other two groups. The Swedish-German bilinguals performed significantly better than the Swedish-Turkish bilinguals on German/Turkish MAIN2 (Baby Birds/Baby Goats), but not on German/Turkish MAIN1 (Cat/Dog) goal comprehension. Although the Swedish-German bilinguals had higher total scores on MAIN1 comprehension in German than the Swedish-Turkish bilinguals in Turkish, both groups thus comprehended goals equally well.

To conclude, the results presented here for comprehension of narrative macrostructure indicate that MAIN2 differentiates better between both age and language groups than MAIN1, but that children do not perform similarly on the two stories of MAIN2, Baby Birds and Baby Goats, whereas performance is identical on the two MAIN1 stories, Cat and Dog.

7.3 Production of macrostructure

This section presents results from the children's production of macrostructural components, analyzed according to the MAIN scoring protocol for macrostructure (Gagarina et al., 2012). The following research questions are asked:

- Are there differences in macrostructure production scores between age groups, and, for Swedish, between language groups?
- Does the use of certain types of components develop differently from others with age, and for monolinguals and bilinguals?
- Which similarities and differences in macrostructural complexity can be seen between the different ages and language groups?
- Do bilinguals perform differently on macrostructure production in their two languages?
- Is performance the same on MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats)?

After describing the scoring and analysis (Section 7.3.1), the results are reported, divided into three parts. First, Section 7.3.2 describes the results for the macrostructure production scores for Swedish, German and Turkish.

Then, Section 7.3.3 looks in detail at the different types of macrostructural components (settings, goals, attempts, outcomes, and internal states) and to what extent they are produced by the children. Finally, Section 7.3.4 analyzes the macrostructural complexity of the children's narratives in terms of different sequences of components (no sequence, attempt-outcome, goal-attempt, goal-outcome, and goal-attempt-outcome, see Section 1.1).

7.3.1 Scoring and analysis

7.3.1.1 Scoring

The MAIN scoring protocol awards points for the successful production of setting (one point each for time and place), and internal state as initiating event, goal, attempt, outcome, and internal state as reaction for three episodes per narrative (one point per component for each episode), yielding a maximum score of 17 points per narrative.[127] Table 7.10 gives an overview with examples of the macrostructural components in Cat (MAIN1).

Table 7.10. Overview of macrostructural components with constructed examples, Cat (MAIN1), setting and episode 1.

Component	MAIN1 – Cat
Setting	Once upon a time… *(time)* …by a lake *(place)*
Episode 1	
IS as IE	A cat saw a butterfly
Goal	The cat wanted to catch the butterfly
Attempt	The cat jumped up…
Outcome	…but landed in the bush
IS as R	The cat was angry

Note. IS = internal state, IE = initiating event, R = reaction.

For Swedish, a total of 165 MAIN1 (Cat/Dog) narratives and 166 MAIN2 (Baby Birds/Baby Goats) narratives were scored.[128] For German, 46 MAIN1 and 46 MAIN2 narratives were scored. For Turkish, 48 MAIN1 and 46 MAIN2 narratives were scored.

In each language, each child received a macrostructure production score on MAIN1 and MAIN2 (out of maximum of 17 points), a combined production score (MAIN1 + MAIN1), and a score on each of the 17 components (counting 'setting' as two separate components; 0 or 1 point per component). For each episode, the child's production was also scored for macrostructural complexity. Here, the focus was on the production of sequences of components within an episode, and the child's production within each episode was

[127] For a description of the different types of components, see Section 1.1.
[128] As described above (Chapter 4), MAIN1 data was missing from one child (BiGer5-14).

classified into one out of four categories: no sequence, attempt-outcome (AO), goal attempt/goal-outcome (GA/GO)[129] or complete episode, i.e. goal-attempt-outcome (GAO).

Scoring of the Swedish and German narratives was carried out by the author. Buket Öztekin scored the Turkish narratives. In addition to the scored examples and scoring principles presented in Gagarina et al. (2012), detailed scoring guidelines (Guidelines for scoring macrostructure in MAIN, version March 2018) were developed in the BiLI-TAS project (Bohnacker, 2013). These scoring guidelines included scoring decisions and the rationale behind them for a more diverse range of cases than the original MAIN guidelines as well as general principles for scoring. They were based on multiple rounds of discussions between native and near-native speakers of Swedish, German and Turkish, who were all linguists and had extensive experience of using MAIN. All scoring was checked thoroughly for consistency against the guidelines.

A macrostructural component had to be realized verbally in a relatively clear manner in order for the child to receive a point. Non-verbal aspects such as pointing or gesturing were not taken into account in the scoring. For the bilingual children, the utterance had to be understandable in the language of testing. In some cases, code-switches influenced the child's score negatively. This was most striking in the German narratives of some Swedish-German bilinguals, whose use of Swedish made certain utterances incomprehensible in a German language context.

A general principle of lenient scoring of reference was used. This was done to avoid that the ability to correctly introduce and refer back to story characters and objects would determine a child's macrostructure production score, i.e. that children who were not able to introduce characters properly would automatically get a lower macrostructure score. The child did thus not need to show adultlike use of reference in order to score on the macrostructural components.[130] Scoring focused on the action (or emotion) described by the child and whether or not this was clear enough. The definition of clear enough here meant that a described action/emotion was understandable and could be linked to a specific character and point in the story (i.e. in the pictures). This meant that the use of relatively specific verbs as well as the inclusion of central agents/patients formed a central aspect in the scoring:

[129] In the analyses, goal-attempt (GA) and goal-outcome (GO) following the MAIN scoring protocol (Gagarina et al., 2012).

[130] If a strict scoring of reference had been used in the scoring of macrostructural components, some children, especially in the younger groups, would have received fewer points. For the monolingual and Swedish-German children, a comparison between lenient scoring and strict scoring was carried out on the Swedish data. If strict scoring had been used, the monolinguals would together have received 33 points less in MAIN1 and 8 points less in MAIN2. The corresponding numbers for the Swedish-German bilinguals were 10 in MAIN1 and 9 in MAIN2. For the data from the Swedish-Turkish children, no such comparison was made, but it is reasonable to assume that numbers would be similar to those of the other groups.

when a child used a verb that was so general that it could refer to a number of situations in the stories, no point could be given for a specific component. In some cases, a verb that could be used to describe different events in the story was produced together with the appropriate agent and/or patient, thereby making it clear which part of the story was being referred to. Such utterances could then be awarded a point. For example, in the Baby Goats story, the Swedish general verb *ta* 'take' can be used to describe both the fox catching the baby goat (Episode 2, outcome) and the bird biting the fox (Episode 3, attempt) (cf. Figure 3.3). A child who produced a rudimentary utterance that only consisted of *ta* is thus not awarded a point. Only by including the correct agent/patient (by using a lexical NP or a pronoun with a clear anaphoric relationship to the correct character) will it be clear enough which component the child is trying to express.

7.3.1.2 Statistical analyses

The macrostructure production scores were analyzed in the same way as the comprehension scores. Correlations with age in months were carried out on combined production scores (i.e. MAIN1 + MAIN2 scores) for each language, and for Swedish for the language groups separately. The bilinguals' combined production scores in their respective languages, broader measures of their narrative production, were also correlated.

In each of the three languages, paired-samples t-tests were run to compare the MAIN1 and MAIN2 production scores, and these scores were also correlated. The scores from the bilinguals' two languages were also compared and correlated. Scores in each language from Cat and Dog and Baby Birds and Baby Goats were compared using independent-samples t-tests. The effect of the order of testing was analyzed for each of the bilinguals' languages. Whenever story or order of testing was significant, it was included in the ANOVA analyses carried out on the production scores from the MAIN1 and MAIN2.

For Swedish macrostructure production scores, two factorial ANOVAs were carried out, one for MAIN1 and one for MAIN2 in order to test effects of language group and age group. Post-hoc tests (Bonferroni) were always used for the variables language group and age group to find out between which groups there were significant differences. To analyze effects of age group on Turkish and German macrostructure production, either factorial or regular ANOVAs with post-hoc tests (Bonferroni) were used, depending on whether the variables story and/or order of testing were significant. Finally, scores from Turkish and German were compared using independent-samples t-tests.

In Swedish, for both MAIN1 and MAIN2, separate independent-samples Kruskal-Wallis tests were run on the different types of component to determine whether there were significant differences between the language groups in the extent to which specific components were produced. Whenever

the test was significant, pairwise comparisons (with adjusted p-values using Bonferroni corrections) were run. The same type of analysis was run with age group as the independent variable.[131] Mann-Whitney U tests were used to compare the bilinguals' production in German and Turkish. The children's production on the two narrative tasks was not compared statistically for types of components.

For macrostructural complexity, the main part of the analyses was carried out on the proportions of different types of sequences. For each language, the production in MAIN1 and MAIN2 was compared using Chi-Square tests. Two logistic regression analyses were run in Swedish and two on the combined data from Turkish and German. The first logistic regression analysis had the proportions of no sequence versus any type of sequence as dependent variable and the second was on the proportion of GAO versus no GAO (including no sequence). In Swedish, both analyses included the variables age group and language group and the interactions between them as predictors. In the analysis of the minority languages, the predictors were the variables age group and language (German vs Turkish) and the interactions between them. The variables age group and language group were coded in the same way as in the analysis of different types of referring expressions (see Section 6.3.2), i.e. using reversed Helmert coding for age group and Helmert coding for language group. Model selection was applied to find the simplest model with the best fit, following the procedure described in Section 6.3.2.

7.3.2 Macrostructure production scores

7.3.2.1 Swedish

Table 7.11 shows the results from the Swedish MAIN1, MAIN2 and combined macrostructure production scores for the three language groups. Even the best-performing children scored far from the maximum. The highest score for a single narrative is 11 points (out of 17) and the highest combined score is 21 points (out of 34). Although mean scores of the monolinguals and the Swedish-German bilinguals are higher than those of the Swedish-Turkish bilinguals, in each language group some children scored very low and others scored relatively high. Two of the narratives which were awarded 11 points, one from MAIN1 and one from MAIN2, as well as narratives with average (6-7 points) and lower-than-average (3-4 points) scores are found in Appendix 1.

[131] As the dependent variables here were proportions, non-parametric tests were used. As there is no standard multivariate non-parametric test, it was not possible to include both language and age group in the same test. This means that no test has been conducted that includes the interactions between language group and age group.

Table 7.11. Swedish macrostructure production scores, MAIN1, MAIN2 and combined, by language group.

	MAIN1	MAIN2	Combined
Monolinguals (N=72)			
Mean (SD)	6.0 (2.1)	6.2 (2.1)	12.2 (3.5)
Range	0 – 11	2 – 10	2 – 18
Swedish-German bilinguals (N=46)			
Mean (SD)	6.4 (2.0)	6.6 (2.4)	13.0 (4.2)
Range	3 – 11	1 – 11	4 – 21
Swedish-Turkish bilinguals (N=48)			
Mean (SD)	4.4 (2.2)	4.4 (2.9)	8.8 (4.4)
Range	0 – 9	0 – 10	0 – 17

Note. Max score for MAIN1 and MAIN2 = 17 points, max score for combined = 34 points.

Before proceeding to analyze the scores of MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) separately, something further should be said about the combined production scores, as these may better reflect the children's broader narrative ability than scores on a single task. There were significant and strong positive correlations between age and combined production score in all three language groups (monolinguals: $r = .645$, $p < .001$, Swedish-German bilinguals: $r = .695$, $p < .001$; Swedish-Turkish bilinguals: $r = .553$, $p < .001$). These relationships between age and the macrostructure production score can be seen clearly in Figure 7.13.

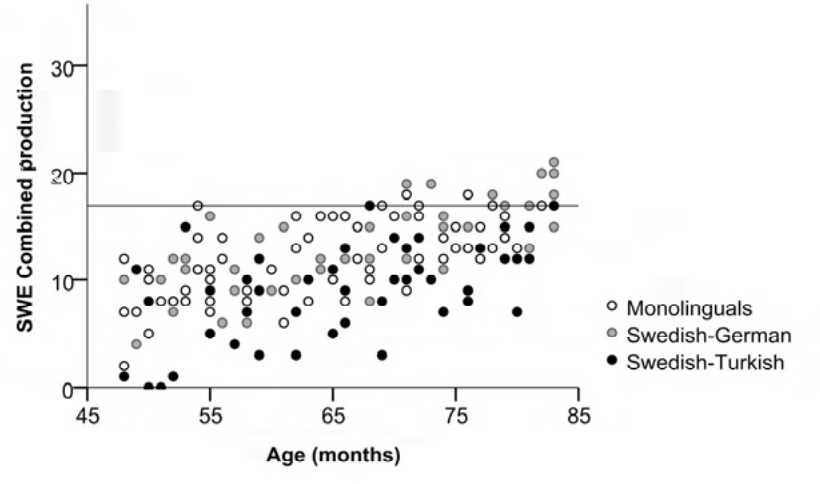

Figure 7.13. Combined (MAIN1 + MAIN2) production of macrostructure, Swedish. Max score = 34 points. A dot may represent more than one individual child. Line indicates 50% (17 points).

The oldest children scored highest, and the lowest scores are all found among the youngest children. These included a number of four-year-olds who scored at or close to zero. However, low-scoring Swedish-Turkish children are found in all age groups. Figure 7.13 also shows that only a small number of older children reached a combined score of at least 50% (17 points) out of the total maximum. The scores of the rest of the children are evenly spread between zero and 50%. Overall, the combined scores from production of macrostructure show that the children produced narratives that contain only some of the macrostructural components included in the MAIN scoring protocol.

Swedish MAIN1 and MAIN2 macrostructure production scores for the different age groups within each language group are shown in Figure 7.14.

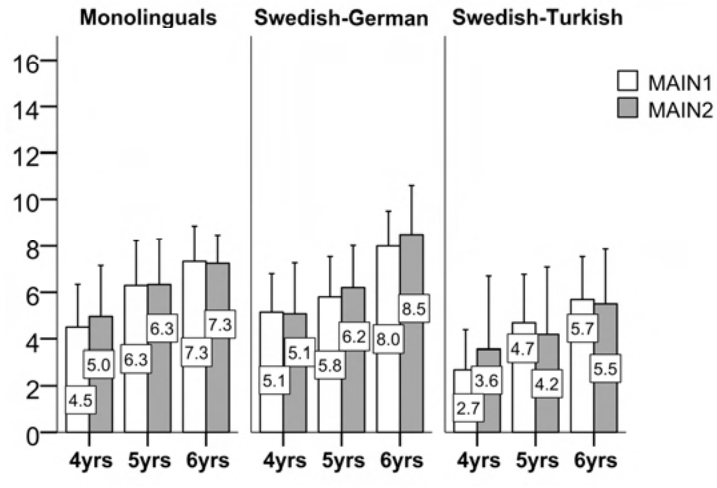

Figure 7.14. Mean macrostructure production scores for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), by language and age group. Max score = 17 points. Error bars show +1SD.

As seen in Figure 7.14, the variation in scores (measured by SDs) was relatively large in all groups, although somewhat smaller for the monolingual and Swedish-German six-year-olds, and in general larger for the Swedish-Turkish bilinguals than for the other language groups. Differences in production scores between MAIN1 and MAIN2 were small in all groups, and the paired-samples t-test showed no significant difference between the two narrative tasks ($t(164) = -.909$, $p = .365$). In contrast to comprehension, the children did not perform better on MAIN1 (Cat/Dog) than on MAIN2 (Baby Birds/Baby Goats) in production.

There was no significant difference between the children's production scores on Cat (M = 5.8, SD = 2.3) and Dog (M = 5.5, SD = 2.3) for MAIN1 ($t(163) = 1.120$, $p = .264$), and also not between Baby Birds (M = 5.9, SD = 2.4) and Baby Goats (M = 5.7, SD = 2.8) for MAIN2 ($t(164) = .469$, $p =$

.639). Contrary to what was the case for the comprehension scores, the children's performance in production is thus not influenced by the differences between Baby Birds and Baby Goats.

As would be expected from Figure 7.14, the factorial Age group x Language group (3x3) ANOVA for MAIN1 showed that there were significant main effects of both language group ($F(2, 156) = 17.770$, $p < .001$, $\eta_p^2 = .186$) and age group ($F(2, 156) = 35.218$, $p < .001$, $\eta_p^2 = .311$) but no significant interaction effect ($F(4, 156) = .848$, $p = .497$, $\eta_p^2 = .021$). The subsequent post-hoc tests showed that there was no difference between the monolinguals and the Swedish-German bilinguals, but that the Swedish-Turkish bilinguals performed significantly lower than the two other groups. There were significant differences between all three age groups; six-year-olds performed better than five-year-olds, who in their turn performed better than four-year-olds. The results from the factorial ANOVA run on the MAIN2 data were identical, with main effects of language group ($F(2, 157) = 13.192$, $p < .001$, $\eta_p^2 = .144$) with significantly lower scores for the Swedish-Turkish children, and of age groups ($F(2, 157) = 17.928$, $p < .001$, $\eta_p^2 = .185$), with significant differences between all three age groups, and no interaction effect ($F(4, 157) = .776$, $p = .542$, $\eta_p^2 = .019$).[132] For production of macrostructure in the Swedish narratives, there are thus clear effects of both age and language group, irrespective of whether Cat/Dog or Baby Birds/Baby Goats is used. In terms of scores, the Swedish-Turkish bilinguals are around one year behind the children in other language groups, i.e. the Swedish-Turkish five-year-olds had comparable scores to monolingual and Swedish-German four-year-olds, and the scores of the Swedish-Turkish six-year-olds resembled those of the five-year-olds in the other groups.

7.3.2.2 Swedish-German bilinguals

In Table 7.12, the results for Swedish and German MAIN1, MAIN2 and combined macrostructure production scores are shown for the Swedish-German bilinguals.

Table 7.12. Macrostructure production, Swedish-German bilinguals (N=46), MAIN1, MAIN2 and combined, by language.

	MAIN1	MAIN2	Combined
Swedish			
Mean (SD)	6.4 (2.0)	6.6 (2.4)	13.0 (4.2)
Range	3 – 11	1 – 11	4 – 21
German			
Mean (SD)	5.3 (2.5)	5.4 (2.7)	10.7 (4.7)
Range	0 – 10	1 – 12	1 – 21

Note. Max score for MAIN1 and MAIN2 = 17 points. Max score for combined = 34 points.

[132] For p-values of all pairwise comparisons for the Swedish MAIN1 and MAIN2 production scores, see Appendix 5, Tables A5.6 and A5.7.

In production, the children performed significantly better in Swedish than in German, both on MAIN1 ($t(44) = 2.847$, $p = .007$) and on MAIN2 ($t(45) = 3.329$, $p = .002$), contrary to in comprehension where there was no difference between the languages. There were significant positive correlations between scores in the two languages both for MAIN1 ($r = .389$, $p = .008$) and MAIN2 ($r = .493$, $p = .001$); children who performed well in one language also tended to do so in the other language. Combined scores were also significantly correlated in the two languages ($r = .480$, $p = .001$), as illustrated by Figure 7.15.

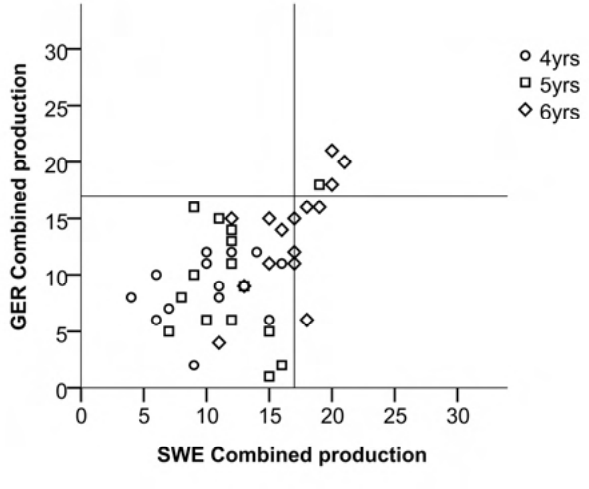

Figure 7.15. Combined (MAIN1 + MAIN2) macrostructure production scores, Swedish and German, Swedish-German bilinguals. Max score = 34 points. A dot may represent more than one individual child. Lines on both axes indicate 50% (= 17 points).

Figure 7.15 shows that most children scored below 50% in both languages. Four children scored above 50% in both languages, and three children scored above 50% only in Swedish. Two of these three children scored just below 50% in German, whereas one child (BiGer6-18) only scored a third of the points in German (6 points) compared to Swedish (18 points). Additionally, two children (BiGer5-01, BiGer5-13) scored close to 50% in Swedish (15 and 16 points, respectively) while scoring close to zero (1 and 2 points only) in German. These three children may have too limited vocabulary and syntax to be able to express the narrative content in an understandable way in German, but do not have any difficulties doing so in Swedish.[133]

[133] The difference between these three children's competencies in Swedish and German is also obvious when comparing their vocabulary production scores (cf. Section 5.2.3). All three scored at least twice as high in Swedish as in German (BiGer5-01 – Swedish: 45, German: 18; BiGer5-13 – Swedish: 50, German: 17; BiGer6-18 – Swedish: 50, German: 25).

There was a significant positive correlation between the combined score in German and age in months (r = .394, p = .007), but this correlation was not as strong as for Swedish (see Section 7.3.2.1). A closer look at the pattern shown in Figure 7.16 reveals that low-scoring children are found in all age groups, but that the highest-scoring children are six-year-olds and older five-year-olds.

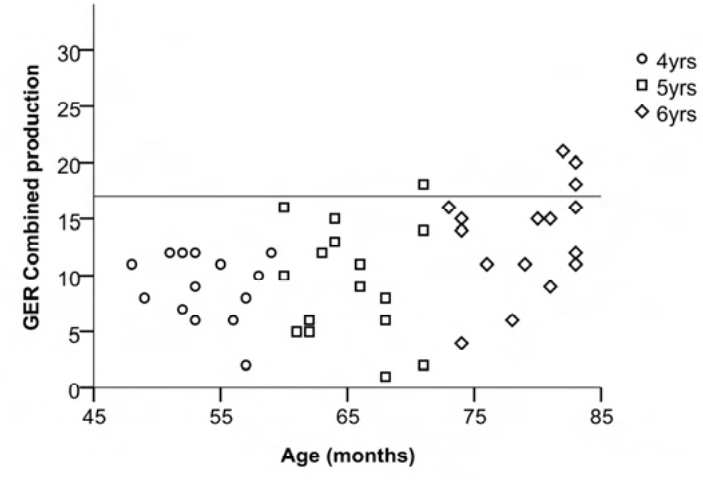

Figure 7.16. German combined (MAIN1 + MAIN2) production scores and age (months), Swedish-German bilinguals. Max score = 34 points. A dot may represent more than one individual child. Line indicates 50% (= 17 points).

In Figure 7.17, MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) macrostructure production scores for Swedish and German are shown by age group. In both languages, production scores for the two different narratives tasks are very similar. In German, just as was the case in Swedish (cf. 7.3.2.1), there was no significant difference between the scores on MAIN1 and MAIN2 (t(45) = -0.210, p = .834).

In German, despite relatively large differences in mean scores between the two stories for both MAIN1 (Cat: M = 5.7, SD = 2.3; Dog: M = 4.9, SD = 2.8) and MAIN2 (Baby Birds: M = 6.0, SD = 2.6; Baby Goats: M = 4.7, SD = 2.6), these differences were not significant (MAIN1: t(44) = 1.051, p = .299; MAIN2: t(44) = 1.604, p = .116). In neither Swedish nor German were there any effects of the order of testing.[134] Thus, whether German was tested first or second did not influence the children's production scores, contrary to the results for narrative comprehension.

[134] Swedish, MAIN1: t(43) = 0.046, p = .964; MAIN2: t(44) = 1.026, p = .310; German, MAIN1: t(44) = 0.696, p = .490; MAIN2: t(44) = 0.498, p = .621.

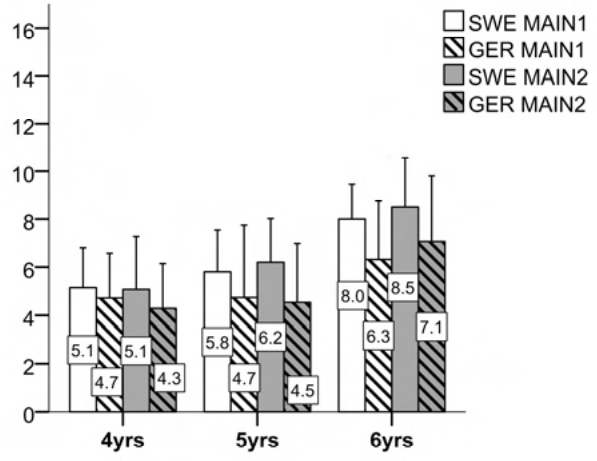

Figure 7.17. Mean macrostructure production scores, Swedish and German MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), Swedish-German bilinguals, by age group. Max score = 17 points. Error bars show +1 SD.

For the German MAIN1, the one-way ANOVA showed no significant effect of age group ($F(2, 44) = 2.046$, $p = .142$). The correlation between MAIN1 production scores and age was also not significant ($r = .281$, $p = .058$). For the German MAIN2 there was a clear significant effect of age group ($F(2, 44) = 6.283$, $p = .004$): the six-year-olds performed better than both younger groups.[135]

To summarize, the Swedish-German bilinguals performed better in Swedish than in German on both MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats). There was no difference between scores on MAIN1 and MAIN2 in German and no effect of order of testing on the score in any of the two languages. In German, there was no age effect for MAIN1 (Cat/Dog), but, on MAIN2 (Baby Birds/Baby Goats), the six-year-olds scored significantly higher than the younger children.

7.3.2.3 Swedish-Turkish bilinguals

In Table 7.13, Swedish and Turkish MAIN1, MAIN2 and combined macrostructure production scores are shown for the Swedish-Turkish bilinguals. As the descriptive statistics in Table 7.13 suggest, there were no differences between the Swedish-Turkish children's scores in Swedish and Turkish, neither for MAIN1 ($t(47) = .844$, $p = .403$), nor for MAIN2 ($t(47) = -.427$, $p = .671$). There was no correlation between scores in the two languages for MAIN1 ($r = .053$, $p = .723$), whereas scores on the Swedish MAIN2 were significantly correlated with the Turkish MAIN2 scores ($r = .314$, $p = .03$).

[135] Exact p-values for the pairwise comparisons can be found in Appendix 5, Table A5.7.

Combined scores in the two languages were also significantly correlated (r = .352, p = .014), as can be seen in Figure 7.18.

Table 7.13. Macrostructure production scores, Swedish-Turkish bilinguals (N=48), MAIN1, MAIN2 and combined, by language.

	MAIN1	MAIN2	Combined
Swedish			
Mean (SD)	4.4 (2.2)	4.4 (2.9)	8.8 (4.4)
Range	0 – 9	0 – 10	0 – 17
Turkish			
Mean (SD)	4.0 (2.2)	4.6 (2.9)	8.6 (4.4)
Range	0 – 10	0 – 9	0 – 16

Note. Max score for MAIN1 and MAIN2 = 17 points. Max score for combined = 34 points.

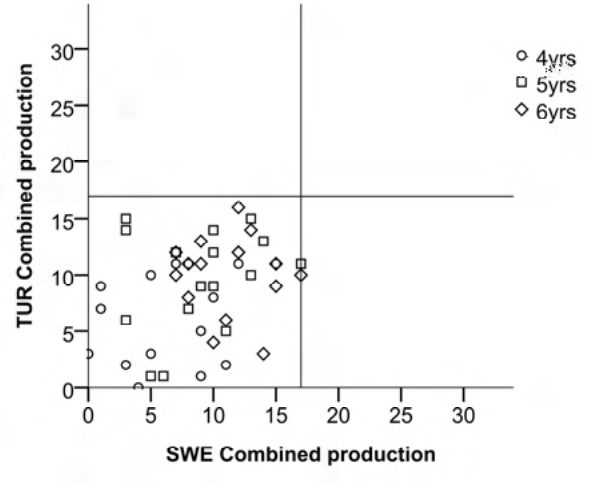

Figure 7.18. Combined (MAIN1 + MAIN2) macrostructure production scores, Swedish and Turkish, Swedish-Turkish bilinguals. Max score = 34 points. A dot may represent more than one individual child. Lines on both axes indicate 50% (= 17 points).

Figure 7.18 shows that all children scored at or below 50% in both languages. In each language, a number of children had very low scores. There was a significant correlation between combined scores in Turkish and age in months (r = .430, p = .002). This correlation was somewhat weaker than the one between combined Swedish scores and age (cf. Section 7.3.2.1), just as was the case for the Swedish-German bilinguals. Looking at scores of individual children in Turkish plotted against the child's age in months, as shown in Figure 7.19, reveals that the children who scored highest in Turkish were five- and six-year-olds and that few of the six-year-olds were among the lowest-scoring children.

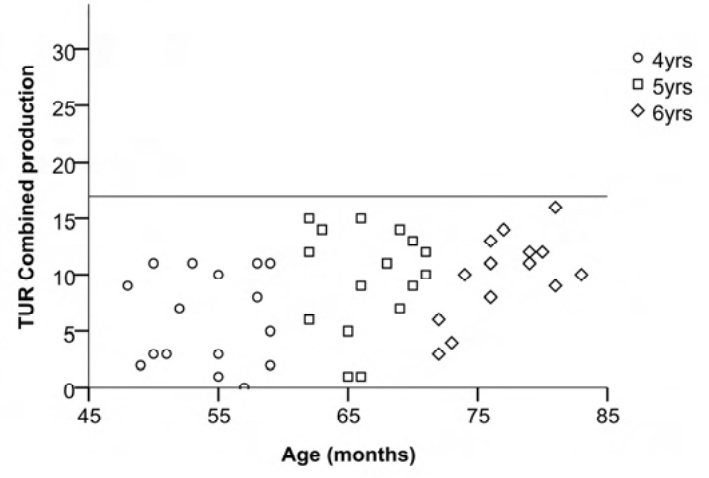

Figure 7.19. Turkish combined (MAIN1 + MAIN2) production scores and age (months), Swedish-Turkish bilinguals. Max score = 34 points. A dot may represent more than one individual child. Line indicates 50% (= 17 points).

In Figure 7.20, macrostructure production scores for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) in Swedish and Turkish are shown by age group. Scores for MAIN1 and MAIN2 are relatively similar in both languages. Although scores on the Turkish MAIN2 were somewhat higher than those of the Turkish MAIN1 in all age groups, this difference was not significant ($t(47) = -1.637$, $p = .108$).

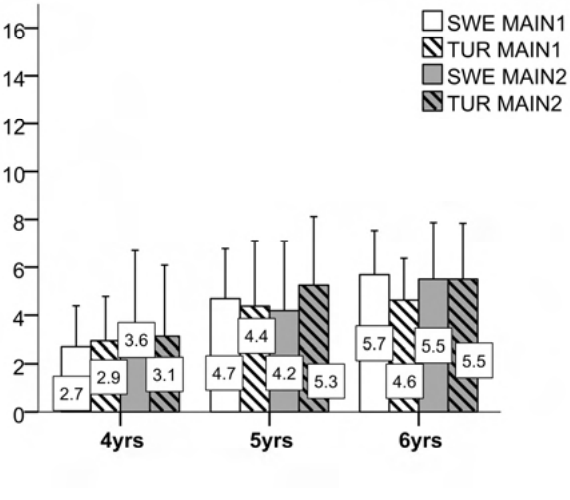

Figure 7.20. Mean macrostructure production scores, Swedish and Turkish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), Swedish-Turkish bilinguals, by age group. Max score = 17 points. Error bars show +1 SD.

In Turkish, there was no difference between the two stories of each task. Scores for Cat (M = 4.0, SD = 2.0) and Dog (M = 4.0, SD = 2.5) were identical ($t(46) = .064$, $p = .949$) and scores for Baby Birds (M = 4.4, SD = 3.2) and Baby Goats (M = 4.8, SD = 2.6) were also highly similar ($t(46) = -.459$, $p = .648$).

There were no effects of order of testing in any of the two languages (Swedish: MAIN1: $t(46) = .836$, $p = .407$, MAIN2: $t(46) = -.497$, $p = .621$; Turkish: MAIN1: $t(46) = 1.505$, $p = .139$, MAIN2: $t(46) = .899$, $p = .374$). Whether or not Turkish or Swedish was tested first did not influence the Swedish-Turkish children's macrostructure production scores.

For the Turkish MAIN1, there was no significant difference between the age groups ($F(2, 46) = 2.865$, $p = .067$), but there was a significant overall effect of age group for the Turkish MAIN2 ($F(2, 46) = 3.641$, $p = .034$). However, the subsequent pairwise comparisons did not reveal any significant differences between any two age groups, although the difference between the four- and the six-year-olds was approaching significance ($p = .054$).[136] There were significant correlations with age (in months) for MAIN1 ($r = .369$, $p = .01$) and MAIN2 ($r = .366$, $p = .011$) in Turkish. Although there are no distinct age group differences in macrostructure scores, there is thus a development with linear age in the Swedish-Turkish bilinguals' minority language Turkish.

To summarize, the Swedish-Turkish bilinguals performed similarly in Swedish and Turkish. There was no difference between MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) in Turkish, and only an overall effect of age group on the MAIN2 scores.

7.3.2.4 German and Turkish

In Table 7.14, macrostructure production scores in the minority languages, German and Turkish are shown.

Table 7.14. Macrostructure production scores, German and Turkish.

	MAIN1	**MAIN2**
Swedish-German bilinguals (German, N=46)		
Mean (SD)	5.3 (2.5)	5.4 (2.7)
Range	0 – 10	1 – 10
Swedish-Turkish bilinguals (Turkish, N=48)		
Mean (SD)	4.0 (2.2)	4.6 (2.9)
Range	0 – 10	0 – 9

Note. Max score = 17 points.

[136] For the p-values of the other pairwise comparisons, see Table A5.6 in Appendix 5.

For MAIN1 (Cat/Dog), the Swedish-German bilinguals performed better in German than the Swedish-Turkish bilinguals did in Turkish ($t(92) = 2.694$, $p = .008$). The Swedish-German bilinguals' scores for MAIN2 (Baby Bird/Baby Goats) were also somewhat higher than those of the Swedish-Turkish bilinguals, but this difference was not large enough to be significant ($t(92) = 1.303$, $p = .196$). Variation in production scores in the groups was similar and relatively large, both in terms of standard deviations and score ranges. The results for production of narrative macrostructure in the minority languages are thus different than for comprehension (recall Section 7.2.2.4), where the Swedish-German bilinguals performed significantly better than the Swedish-Turkish bilinguals on both MAIN1 and MAIN2.

7.3.2 Different types of macrostructural components

In this section, results from macrostructure production are reported for each of the six different types of macrostructural components (settings, internal states as initiating events (ISTs as IE), goals, attempts, outcomes and internal states as reaction (ISTs as R), cf. Section 1.1)) included in the MAIN macrostructure production score. The results are reported as *percentage overtly realized* (i.e. produced) for each type of component (which mean that for each component, the maximum is 100%) in Swedish, German and Turkish. Results for the language and age groups on the different components by episode (i.e. for the 17 components separately, including results for the two setting components time and place) can be found in Appendix 5, Tables A5.8-A5.11.

7.3.2.1 Swedish

In Figure 7.21, the percentages of each type of component that were overtly realized by the children in their Swedish MAIN1 (Cat/Dog) narratives are shown by language group for all age groups combined.

Figure 7.21 shows clear differences in all three language groups in the extent to which the different components are produced by the children. Internal states as reaction were almost never included. Settings, internal states as initiating events and goals were produced more often, but were still relatively infrequent in the children's narratives. By far the most frequent components were attempts and outcomes, with the latter being more frequent.

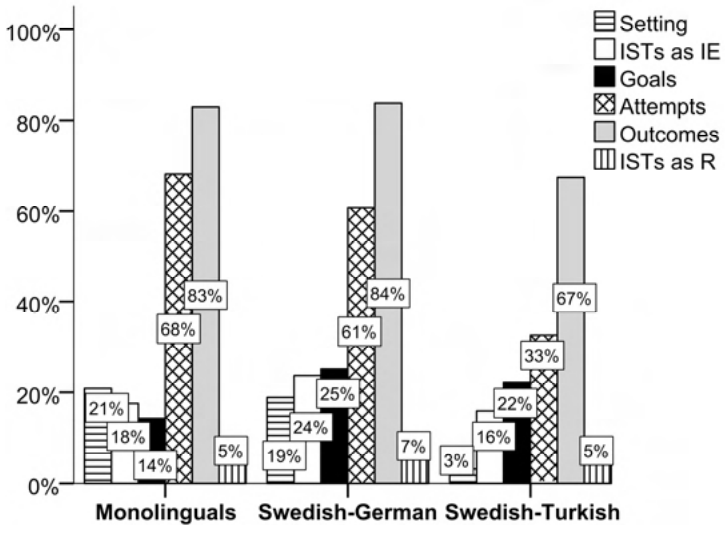

Figure 7.21. Percentage (%) overtly realized components by type, Swedish MAIN1 (Cat/Dog), by language group (all age groups). Max per type of component = 100%.

In (21), (22) and (23), examples of MAIN1 ISTs are shown (underlined). Examples of goals, attempts and outcomes are found in Section 7.3.3.[137]

(21) och då blir katten hungrig på dom där små fiskarna som ligger i hinken
'and then the cat becomes hungry for those small fish lying in the bucket.'
(MoSwe4-14, 4;6 – Cat, IST as IE, Episode 3)

(22) och sen så hoppa(de) katten ner i taggbusken så den fick ont
'And then the cat jumped down into the thorn bush so it got pain.'
(MoSwe5-10, 5;6 – Cat, IST as R, Episode 1)

(23) och när mannen hitta(de) ballongen så blev han glad
'And when the man found the balloon, so he became happy.'
(BiTur6-02, 6;9 – Dog, IST as R, Episode 2)

As would be expected from Figure 7.21, there were no significant group differences on MAIN1 for the two types of ISTs (ISTs as IE: $H(2) = 1.791$, $p = .408$; ISTs as R: $H(2) = .630$, p = .730). The Kruskal-Wallis tests for the other four components were all significant (Settings: $H(2) = 12.221$, p = .002; Goals: $H(2) = 7.805$, $p = .02$; Attempts: $H(2) = 37.178$, $p < .001$; Outcomes: $H(2) = 12.661$, $p = .002$). The Swedish-Turkish bilinguals produced a lower proportion of settings, attempts and outcomes than the two other groups. The monolinguals produced a lower proportion of goals than the Swedish-German bilinguals. The difference in the proportions of goals be-

[137] In all examples shown in this chapter, pauses, hesitations, and repetitions have been removed for increased readability.

tween the monolinguals and the Swedish-Turkish bilinguals, although fairly large (cf. Figure 7.21), did not reach significance.[138]

Figure 7.22 shows the results from Figure 7.21 separately for the three age groups within each language group for the Swedish MAIN1 (Cat/Dog).

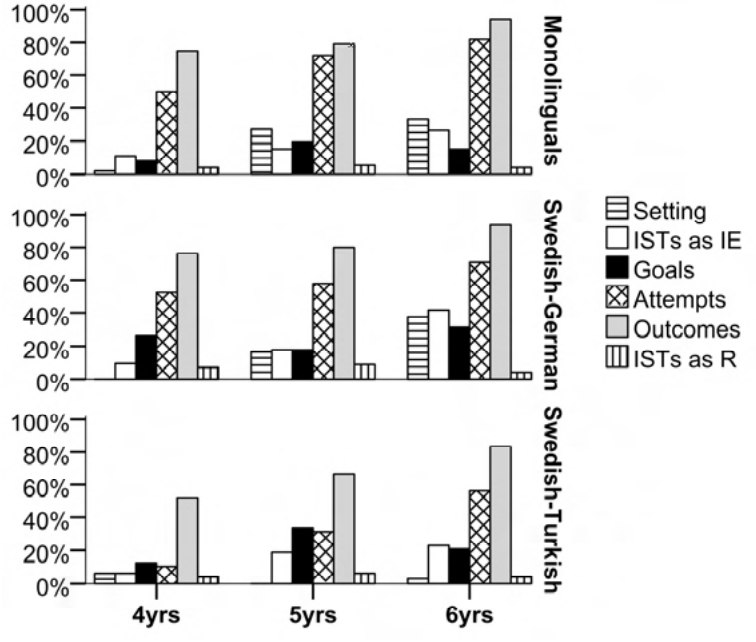

Figure 7.22. Percentage (%) overtly realized components by type, Swedish MAIN1 (Cat/Dog), by language and age group. Max per type of component = 100%.

Age patterns are fairly similar across language groups. In all age groups, outcomes are the most frequent component, being produced by the majority of all children in MAIN1. In all groups except the Swedish-Turkish four- and five-year-olds, proportions of attempts are also at least 50%.

For MAIN1, there were significant differences between the age groups for settings ($H(2) = 20.301$, $p < .001$), IST as IE ($H(2) = 17.761$, $p < .001$), attempts ($H(2) = 30.867$, $p < .001$) and outcomes ($H(2) = 18.651$, $p < .001$), whereas goals ($H(2) = 4.291$, $p = .117$) and ISTs as R ($H(2) = 1.898$, $p = .387$) did not develop with age. The six-year-olds produced more settings than the four-year-olds. The six-year-olds also produced significantly more ISTs as IE than both the four-year-olds and five-year-olds, but there was no difference between the younger groups. The same development with age was also found for outcomes. For attempts, there were clear significant differences between all three groups; the six-year-olds produced more attempts

[138] P-values for all pairwise comparisons can be found in Appendix 5, Table A5.12.

than the five-year-olds who in their turn produced more than the four-year-olds.[139]

Moving on to the results for MAIN2, Figure 7.23 shows the percentages of each type of component that were overtly realized by the children in the Swedish MAIN2 (Baby Birds/Baby Goats) narratives, by language group.

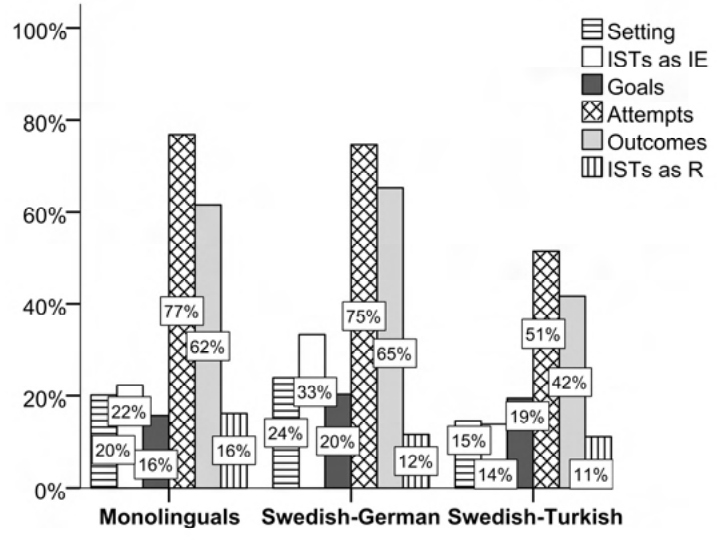

Figure 7.23. Percentage overtly realized components by type, Swedish MAIN2 (Baby Birds/Baby Goats), by language group (all age groups). Max per type of component = 100%.

For MAIN2, results were somewhat different than for MAIN1. First, attempts were more common than outcomes, and internal states were more frequently produced. In MAIN2, contrary to MAIN1, there were no group differences for settings ($H(2) = 3.130$, $p = .209$) and goals ($H(2) = 1.978$, $p = .372$). For ISTs as IE, the Swedish-German bilinguals produced a higher proportion than the Swedish-Turkish bilinguals, but there were no differences between either bilingual group and the monolinguals ($H(2) = 8.780$, $p = .012$). There was no difference for ISTs as R ($H(2) = 1.785$, $p = .410$). Additionally, there were significant group differences in attempts ($H(2) = 14.028$, $p = .001$) and outcomes ($H(2) = 13.101$, $p = .001$), with the Swedish-Turkish bilinguals again producing fewer attempts and outcomes than the Swedish-German bilinguals and the monolinguals. For neither component there was a difference between the monolinguals and the Swedish-German bilinguals.[140] In (24), an example of an IST from MAIN2 is shown.[141]

[139] All pairwise comparisons are shown in Appendix 5, Table A5.13.
[140] In Appendix 5, Table A5.14, all pairwise comparisons are shown.
[141] Note that this example also includes a goal (*vill äta upp fåglarna* 'wants to eat the birds').

(24) men nedanför trädet är en hungrig katt som vill äta upp fåglarna
'But below the tree is a hungry cat who wants to the birds.'
(BiGer6-06, 6;11 – Baby Birds, IST as IE, Episode 2)

In Figure 7.24, Swedish MAIN2 (Baby Birds/Baby Goats) results for the six types of components are shown by age group for the three language groups.

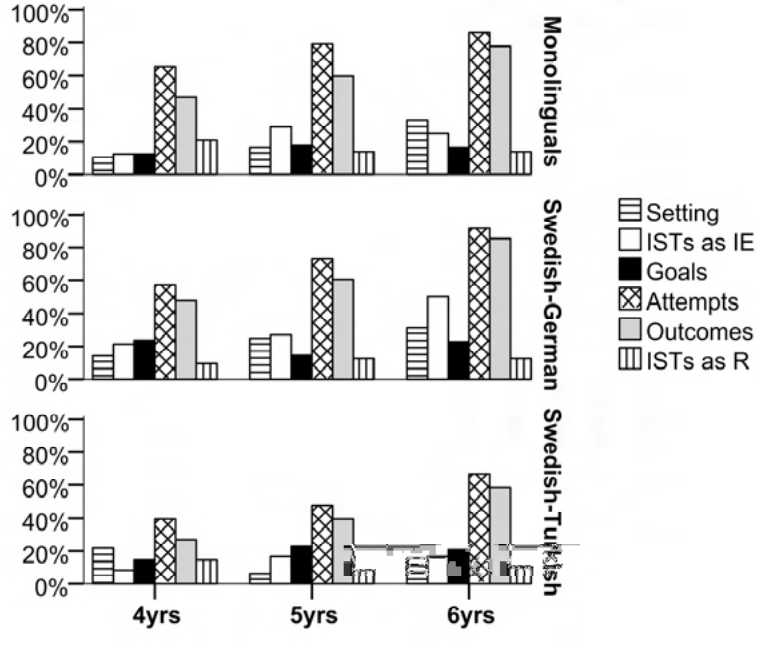

Figure 7.24. Percentage overtly realized components by type, Swedish MAIN2 (Baby Birds/Baby Goats), by language and age group. Max per type of component = 100%.

In MAIN2, attempts and outcomes were the most commonly produced components in all age groups, with attempts being somewhat more frequent. For MAIN2, just as for MAIN1, goals and internal states as reaction were produced to a similar extent by all age groups (goals: $H(2) = 1.097$, $p = .578$; IST as R: $H(2) = 1.84$, $p = .399$). There were significant effects of age for settings ($H(2) = 6.725$, $p = .035$), ISTs as IE ($H(2) = 10.335$, $p = .006$), attempts ($H(2) = 18.397$, $p < .001$) and outcomes ($H(2) = 27.866$, $p < .001$). The pairwise comparisons showed no difference between specific age groups for settings. For ISTs as IE, there was only a difference between the six-year-olds and the four-year-olds. The six-year-olds produced a higher percentage of attempts than did the four-year-olds. For outcomes, there were clear differences between the six-year-olds and both younger groups, but no difference between the two younger groups.[142]

[142] All pairwise comparisons are shown in Appendix 5, Table A5.15.

To summarize, in neither MAIN1 nor MAIN2, there was any age development in the use of internal states as reaction and goals; these components were infrequently used by all groups (especially in MAIN1). For settings, internal states as initiating events, attempts and outcomes, there was development with age, although the results varied somewhat between components and narrative tasks. Although attempts and outcomes were already used to a large extent by the youngest children (with the exception of the Swedish-Turkish four- and five-year-olds), the clearest age development was found for these two components.

7.3.2.2 German and Turkish

The only statistically significant difference in overall production scores between specific age groups in the German and Turkish data was between the Swedish-German six-year-olds and the younger children in MAIN2. Therefore, no analyses were carried out on age groups in German and Turkish for the six different types of components.

In Figure 7.25, the percentages of each type of component overtly realized by the children in their German and Turkish MAIN1 (Cat/Dog) narratives are shown by language for all age groups combined.

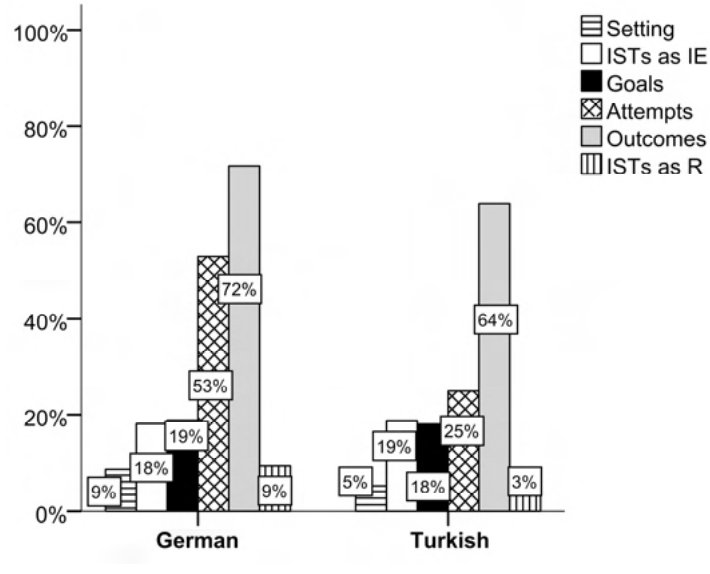

Figure 7.25. Percentage overtly realized components by type, German and Turkish MAIN1 (Cat/Dog), by language (all age groups). Max per type of component = 100%.

Patterns in the two minority languages, German and Turkish are similar for most components, with low proportions of settings and internal states as

reaction and somewhat higher proportions of internal states as initiating event and goals. In both groups, outcomes are by far the most frequently produced component. The only significant difference between German and Turkish was found for attempts ($U = 577.5$, $p < .001$); all other comparisons were non-significant.[143]

In Figure 7.26, results are shown for the different types of components in the German and Turkish MAIN2 (Baby Birds/Baby Goats).

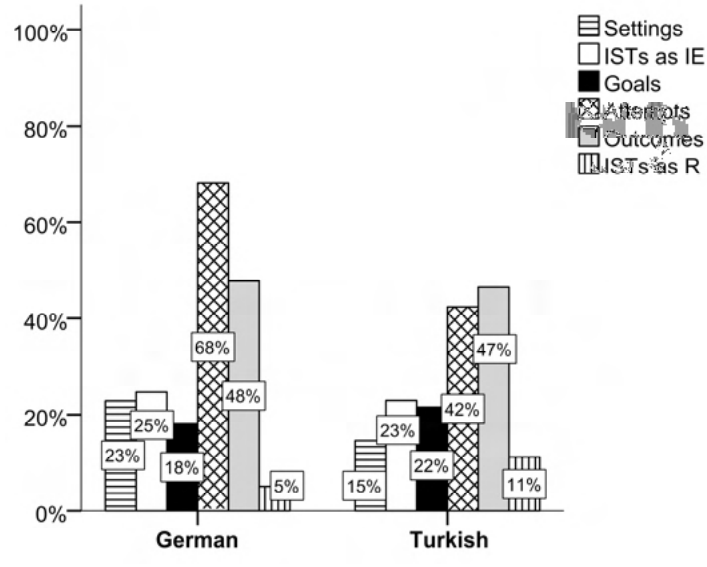

Figure 7.26. Percentage overtly realized components by type, German and Turkish MAIN2 (Baby Birds/Baby Goats), by language (all age groups). Max per type of component = 100%.

In both German and Turkish, just as in Swedish, settings were more common in MAIN2 than in MAIN1, whereas the proportions for both types of ISTs and goals were similar in MAIN1 and MAIN2. For both languages, there were more attempts and fewer outcomes in MAIN2 than in MAIN1. The Mann-Whitney U tests showed the German and Swedish data only differed for attempts and ISTs as R. In German, there were a higher proportion of attempts ($U = 668.5$, $p = .001$), and in Turkish, a higher proportion of ISTs as R was found ($U = 1\,304$, $p = .042$). The comparisons for the other components did not reach significance.[144] In (25), an example of an IST as R in the Turkish data is shown.[145]

[143] Settings: $U = 1{,}027$, $p = .330$; ISTs as IE: $U = 1{,}091$, $p = .911$; goals: $U = 1{,}052$, $p = .657$; outcomes: $U = 940.5$, $p = .192$; ISTs as R: $U = 950$, $p = .075$.
[144] Settings: $U = 956$, $p = .178$; ISTs as IE: $U = 1{,}109$, $p = .967$; goals: $U = 1{,}207$, $p = .387$; outcomes: $U = 1{,}089.5$, $p = .909$.
[145] The symbol / marks the boundary between utterances.

(25) sonra kuş kurdu kovalıyor / sonra bütün koyun mutlu oldu
'And then the bird chases the wolf away / and then all the lamb became happy.'
(BiTur5-09, 5;2 – Baby Goats, IST as R, Episode 3)

7.3.3 Macrostructural complexity

This section reports results for macrostructural complexity, i.e. the use of different sequences of macrostructural components: attempt-outcome (AO), goal-attempt/goal-outcome (GA/GO) and full episodic structures or goal-attempt-outcome (GAO) sequences. The data here is the total number of *opportunities* for the children to produce a sequence of components. Each episode in a narrative is an opportunity. Each opportunity (i.e. each episode) was coded according to the type of sequence the child produced: no sequence, AO, GA/GO or GAO. The main part of the analysis was then run on proportions of different types of sequences (out of all opportunities). To give an example of how the number of opportunities was calculated: in each monolingual age group, there are 24 children, each producing one MAIN1 narrative containing three opportunities to produce a sequence of macrostructural components. This yields 72 possible sequences for each narrative task in each monolingual age group, and a total of 432 possible sequences for the monolinguals as a whole. In Table 7.15, the total numbers of possible sequences are shown.

Table 7.15. Macrostructural complexity, total number of opportunities to produce a sequence (= number of episodes), by language and language group.

	Swedish	**German**	**Turkish**
Monolinguals (N=72)	432		
Swedish-German bilinguals (N=46)	273	276	
Swedish-Turkish bilinguals (N=48)	288		288
Total	**993**	**276**	**288**

7.3.3.1 Swedish

Figure 7.27 shows the distribution of different types of macrostructural sequences (no sequence, AO, GA/GO, GAO) out of all possible sequences (100%) for the Swedish MAIN1 and MAIN2, by language and age group.[146]

[146] An analysis at the level of the different stories (e.g. Cat vs Dog) and episodes is too detailed for the current study, but constitutes one possible way to extend the analysis carried out here.

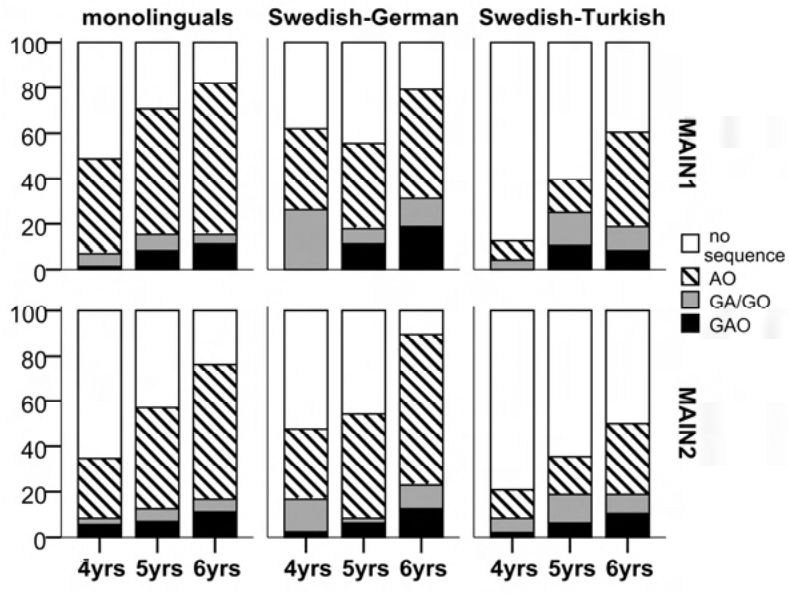

Figure 7.27. Distribution of different types of macrostructural sequences in percentage out of all possible sequences, Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), by age and language group. AO = attempt-outcome, GA = goal-attempt, GO = goal-outcome, GAO = goal-attempt-outcome (complete episode).

A number of interesting observations can be made from the results presented in Figure 7.27. First, when the children produced a sequence it was most commonly AO (see (26) and (27)). The types of sequence thought to be more complex, GA/GO (see (28) and (29)) and GAO (see (30)–(32)), were much rarer. For example, in MAIN2, the monolingual six-year-olds produced 59.7% AO, but only 5.6% GA/GO-sequences and 11.1% GAO-sequences. This is linked to the fact that goals were only infrequently produced by the children (cf. Section 7.3.2.1).

Second, results for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) were similar. This impression is confirmed by a Chi-Square test; there were no significant differences in child performance between the two narrative tasks for macrostructural complexity ($\chi^2(3, N = 993) = 4.063$, $p = .255$). In the subsequent logistic regression analyses, data from MAIN1 and MAIN2 have therefore been combined.

Third, there seems to be an age development both in how often the children produce any sequence at all and in how frequently the different types of sequences are produced. For example, in MAIN1, the monolingual four-year-olds failed to produce any type of sequence in 51.4% of the cases, whereas the six-year-olds only did so in 18.1% of the cases. Fourth, the Swedish-Turkish bilinguals produced much higher proportions of 'no sequence' and much lower proportions of AO-sequences than the children in

the other two groups. The pattern is especially pronounced for the Swedish-Turkish four-year-olds who did not produce any sequence in 87.5% of the cases and only produced 8.3% AO-sequences. This is not surprising given that the Swedish-Turkish bilinguals produced attempts and outcomes to a much lower extent than the children in the other groups (cf. Section 7.3.2.1). In the following, the focus is on comparing the production of any type of sequence to no sequence and on comparing the production of complete episodes, i.e. GAO-sequences, to all types of incomplete episodes (including no sequence, as well as AO- and GA/GO-sequences).

Examples (26)–(32) show AO, GA/GO and GAO sequences in the Swedish data.

(26) han klättra(de) upp och tog ner ballongen
'He climbed up and took down the balloon.'
(MoSwe4-11, 4;3 – Dog, AO-sequence Episode 2)

(27) mamma flög iväg och hämtade mat / och mamma gede mask till ungarna
'Mother flew away and got food / and mother gived worms to the children.'
(BiTur4-26, 4;1 – Baby Birds, AO-sequence Episode 1)

(28) en katt vill jaga en fjäril / och fjärilen flyger iväg
'A cat wants to chase a butterfly / and the butterfly flies away.'
(BiGer4-06, 4;5 – Cat, GO-sequence Episode 1)

(29) och sen vill katten äta dom / och nu kom katten, den klättrade upp
'And then the cat wants to eat them / and now the cat came, it climbed up.'
(BiTur4-02, 4;11 – Baby Birds, GA-sequence Episode 2)

(30) alltså en katt ser en fjäril som sitter på en buske med taggar / och katten hoppar upp och vill ta den, men ramlar istället i taggarna
'So a cat sees a butterfly that sits on a bush with thorns / and the cat jumps up and wants to take it, but instead falls in the thorns.'
(BiGer6-15, 6;1 – Cat, GAO-sequence Episode 1)[147]

(31) och då flög deras mamma iväg för att hämta mat / och sen kom mamman tillbaka med mat, med mask
'And then their mother flew away to get food / and then the mother came back with food, with worms.'
(MoSwe6-03, 6;5 – Baby Birds, GAO-sequence Episode 1)

[147] Note that this example includes an IST as IE (*en katt ser en fjäril* 'a cat sees a butterfly'), in addition to the goal (*vill ta den* 'wants to take it'). Producing both these components in the same episode is very rare among the children, especially with a full GAO-sequence – only 18 episodes with IST as IE + GAO were found in the Swedish MAIN1 and MAIN2 narratives (Monolinguals: 9 episodes, Swedish-German bilinguals: 6 episodes, Swedish-Turkish bilinguals: 3 episodes). This corresponds to only 1.8% of all episodes in the Swedish data (18/993, cf. Table 7.15). In only 2 cases (0.2%, 2/993), one in MAIN1 (MoSwe6-02, see Appendix 1) and one in MAIN2 (BiGer6-09), did a child produce all five macrostructural components (IST as IE, Goal, Attempt, Outcome, IST as R) in the same episode.

(32) sen villde katten äta upp fåglarna / sen katten klättrade upp till trädet, sen tog katten en fågel
'Then the cat wanted to eat the birds / then the cat climbed up to the tree, then the cat took a bird.'
(BiTur6-08, 6;7 – Baby Birds, GAO-sequence Episode 2)

Here we first look at group differences in whether a sequence is produced or not, as producing any type of sequence is clearly better than producing no sequence at all. Table 7.16 shows the summary of the final logistic regression model for sequence versus no sequence for the Swedish data. There were highly significant main effects of age, both for six-year-olds versus the younger groups and for four- versus five-year-olds, with the older groups producing a higher proportion of sequences than the younger groups. There were also significant differences between monolinguals and bilinguals, with the monolinguals producing sequences to a higher extent, and between the Swedish-German bilinguals and the Swedish-Turkish bilinguals, with the former producing a higher proportion of sequences than the latter.[148]

Table 7.16. Summary of logistic regression model 7.1: Macrostructural complexity, sequence versus no sequence, Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) combined.

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	.192	.071	7.335	.007**
Age (1): 6 vs 4 & 5	1.292	.154	70.430	< .001***
Age (2): 5 vs 4	.687	.164	17.493	< .001***
Language group (1): monolinguals vs bilinguals	.502	.140	12.927	< .001***
Language group (2): Swedish-German vs Swedish-Turkish	1.275	.187	46.647	< .001***
Model evaluation				
R^2 (Nagelkerke)	.192			
-2 Log likelihood	1211.871			

Note. ** = p < .01, *** = p < .001. The second value of each predictor is the reference level for that predictor.

The results shown in Figure 7.27 indicate that the Swedish-German bilinguals and the monolinguals performed similarly and the Swedish-Turkish bilinguals differently. Therefore, a second logistic regression model was run in which the variable language group was recoded so that one binary predictor compared the Swedish-Turkish bilinguals with the Swedish-German bilinguals and the monolinguals, and the second predictor compared the Swedish-German bilinguals and the monolinguals. This second model showed a clear significant difference between the Swedish-Turkish bilinguals, who

[148] No interaction effects were significant and adding interactions to the model did not improve model fit ($\Delta r^2 = .009$, p = .092).

produced a lower proportion of sequences, and the other two groups (β = -1.207, *SE* = .156, z = 60.040, p < .001), but not between the Swedish-German bilinguals and the monolinguals (β = .135, *SE* = .169, z = .642, p = .423).[149]

Although analyses at the group level illuminate general trends in the data, it is also important to say something about the performance of the individual children. Table 7.17 shows the percentage of the children who do and do not produce any sequence in each age group.

Table 7.17. Swedish sequences, percentage (%) children producing at least one sequence versus those who do not, by language group and narrative task.

	Monolinguals (N=72)		**Swedish-German (N=46)**		**Swedish-Turkish (N=48)**	
	MAIN1	MAIN2	MAIN1	MAIN2	MAIN1	MAIN2
≥1 sequence	67.1	56.0	65.9	64.5	37.5	35.4
no sequence	32.9	44.0	34.1	35.5	62.5	64.6

There were significant differences between the language groups in the proportion of children producing a sequence, for both MAIN1 (χ^2(2, N = 495) = 35.751, p < .001) and MAIN2 (χ^2(2, N = 498) = 25.877, p < .001). The post-hoc tests (Bonferroni corrected) showed a significant difference between the Swedish-Turkish group, in which a lower proportion of children produced any type of sequence, and the two other groups. In fact, percentages for the Swedish-Turkish children were close to the opposite of the other two groups.

Next, GAO-sequences were analyzed. Table 7.18 shows the final logistic regression model on the proportion of GAO-sequences produced.

Table 7.18. Summary of logistic regression model 7.2: Macrostructural complexity, GAO-sequence versus other (including no sequence), Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) combined.

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	-2.753	.157	306.924	< .001***
Age (1): 6 vs 4 & 5	1.116	.274	16.599	< .001***
Age (2): 5 vs 4	1.381	.432	10.236	.001**
Language group (1): monolinguals vs bilinguals	.014	.248	.003	.955
Language group (2): Swedish-German vs Swedish-Turkish	.345	.328	1.106	.293
Model evaluation				
R^2 (Nagelkerke)	.066			
-2 Log likelihood	499.128			

Note. ** = p < .01, *** = p < .001. The second value of each predictor is the reference level for that predictor.

[149] The full model summary can be found in Appendix 5, Table A5.16.

The model shows a clear main effect of age: the six-year-olds produced a higher proportion of GAOs than the younger groups, and the five-year-olds did so to a larger extent than the four-year-olds. There were no significant effects of language group: no difference between monolinguals and bilinguals and no difference between the two bilingual groups.[150] Table 7.19 shows the percentage of GAOs out of all possible sequences by language and age group.

Table 7.19. Swedish sequences, percentage (%) GAOs out of all possible sequences, by language group and narrative task.

	Monolinguals (N=72)		**Swedish-German (N=46)**		**Swedish-Turkish (N=48)**	
	MAIN1	MAIN2	MAIN1	MAIN2	MAIN1	MAIN2
4-year-olds	1.4	5.6	0.0	2.4	0.0	2.1
5-year-olds	8.3	6.9	11.1	6.3	10.4	6.3
6-year-olds	11.1	11.1	18.8	12.5	8.3	10.4

In addition to the results for percentages of GAO-sequences, looking at the performance of individual children is also important. Table 7.20 shows the percentage of children in each language group who produced at least one complete episode, i.e. GAO-sequence, versus those who produced no GAO, for MAIN1 and MAIN2.

Table 7.20. Swedish GAO-sequences, percentage (%) children producing at least one GAO-sequence, by language group and narrative task.

	Monolinguals (N=72)		**Swedish-German (N=46)**		**Swedish-Turkish (N=48)**	
	MAIN1	MAIN2	MAIN1	MAIN2	MAIN1	MAIN2
≥1 GAO	16.7	19.4	31.1	21.7	14.6	16.7
no GAO	83.3	80.6	68.9	78.3	85.4	83.3

As is shown in Table 7.20, only relatively few of the children in the three language groups produced at least one GAO-sequence. There were no significant differences between the language groups in the proportion of children producing a GAO, neither for MAIN1 ($\chi^2(2, N = 165) = 4.852$, $p = .088$) nor for MAIN2 ($\chi^2(2, N = 166) = .391$, $p = .823$). The results for GAO versus no GAO were thus different from the results for sequence versus no sequence.

7.3.3.2 German and Turkish

In (33) – (38), examples of AO, GA/GO and GAO sequences in the German and Turkish data are shown.

[150] No interaction effects were significant and adding interactions to the model did not improve model fit ($\Delta r^2 = .008$, p = .560).

(33) fare kaçıyor, köpek de onu yakalamaya çalışıyor / sonra ağaca çarptı köpek‘
‘The mouse runs away and the dog tries to catch it / and then the dog hit the tree.’ (BiTur4-05, 4;5 – Dog, AO-sequence, Episode 1)
(34) da biss die Krähe den Fuchs in Schwanz und jagt ihn weg
‘There the crow bit the fox in the tail and chases him away.’
(BiGer6-08, 6;11 – Baby Goats, AO-sequence, Episode 3)

(35) die Katze klettert hoch und will das Vogelbaby schnappen
’The cat climbs up and wants to grab the baby bird.’
(BiGer4-10, 4;1 – Baby Birds, GA-sequence, Episode 2)
(36) o, kuş yavrularını almak istiyo(r)du, / sonra kedi ağaca tırmanıyo(r)du
‘It wanted to take the baby birds / and then the cat was climbing the tree.’
(BiTur5-05, 5;11 – Baby Birds, GA-sequence, Episode 2)

(37) köpek gelmiş, fareyi görmüş / yakalamak istemiş / koşmuş fareye köpek sonra / sonra köpek kafasını çarpmış ağaca
‘The dog came and saw the mouse / wanted to catch (it) / and the dog ran to the mouse / and then the dog hit its head on the tree.’
(BiTur5-19, 5;6 – Dog, GAO-sequence, Episode 1)[151]
(38) der will die Ziegenkinder essen / und jetzt rennt der Fuchs hinter der her und der Fuchs hält auch das ein Bein von dem fest
‘He wants to eat the baby goats / and now the fox runs after her and the fox also holds one of its legs.‘
(BiGer4-09, 4;10 – Baby Goats, GAO-sequence, Episode 2)

Figure 7.28 shows the distribution of different types of macrostructural sequences (no sequence, AO, GA/GO, GAO) out of all possible sequences (100%) for the German and Turkish MAIN1 and MAIN2, by language and age group. When looking at the results presented in Figure 7.28, a number of observations can be made. First, the distribution of different types of sequences is similar in the two narrative tasks. A Chi-square test showed that there were no differences between the types of sequences produced in MAIN1 and MAIN2 ($\chi^2(3, N = 564) = .906$, $p = .824$). For that reason, data from MAIN1 and MAIN2 have been combined in the subsequent logistic regression analyses. Second, the Swedish-German bilinguals had a lower proportion of no sequence than the Swedish-Turkish bilinguals. This difference seems to be mainly due to a higher proportion of AO-sequences in the German data. Third, for the older age groups in both language groups, the proportions of no sequence is lower and the proportion of AO- and GA/GO-sequences is higher, whereas the pattern for GAO-sequences is less clear.

[151] This example includes both an IST as IE (*fareyi görmüş* ‘saw the mouse’) and a GAO. In the German and Turkish narratives, just as in Swedish, few such episodes were found, corresponding to only 1.1% (3/276) and 1.4% (4/288) of the episodes in German and Turkish, respectively. Only one case where a child included all components in the same episode was found in the minority language data (BiGer6-02, Baby Birds, Episode 2).

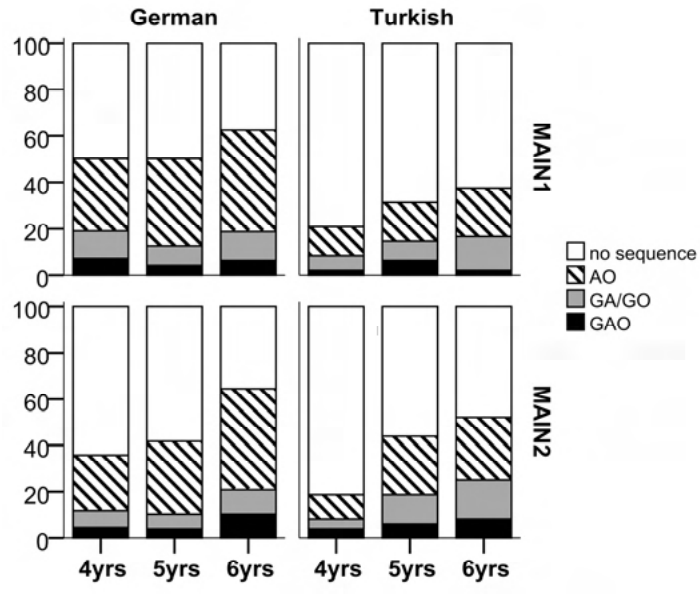

Figure 7.28. Distribution of different types of macrostructural sequences in percentage out of all possible sequences, German and Turkish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), by age and language. AO = attempt-outcome, GA = goal-attempt, GO = goal-outcome, GAO = goal-attempt-outcome (complete episode).

We now analyze the proportions of sequence versus no sequence in the German and Turkish data. Table 7.21 shows the summary of the final logistic regression model for sequence versus no sequence for the German and Turkish data. A significantly higher proportion of sequences was produced in German, and there were clear effects of age. The six-year-olds produced a higher proportion of sequences than the younger groups, and the five-year-olds performed better than the four-year-olds.[152]

[152] No interaction effects were significant and adding interactions to the model did not improve model fit ($\Delta r^2 = .006$, $p = .221$).

Table 7.21. Summary of logistic regression model 7.3: Macrostructural complexity, sequence versus no sequence, German and Turkish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) combined.

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	-.688	.127	29.270	< .001***
Age (1): 6 vs 4 & 5	.757	.184	16.863	< .001***
Age (2): 5 vs 4	.475	.221	4.596	.032*
Language: German vs Turkish	.712	.177	16.252	< .001***
Model evaluation				
R^2 (Nagelkerke)	.088			
-2 Log likelihood	730.698			

Note. * = p < .05, *** = p < .001. The second value of each predictor is the reference level for that predictor.

Table 7.22 shows the percentage of the children who do and do not produce any sequence in each age group for the German narratives of the Swedish-German bilinguals and the Turkish narratives of the Swedish-Turkish bilinguals. Chi-square tests showed that there were significant differences between the language groups for both MAIN narratives (MAIN1: $\chi^2(1, N = 94) = 4.747$, $p = .029$; MAIN2: $\chi^2(1, N = 94) = 3.909$, $p = .048$). In both minority language narratives, a significantly larger proportion of the Swedish-German bilinguals produced at least one sequence.

Table 7.22. German and Turkish sequences, percentage (%) children producing at least one sequence versus those who do not, by language group and narrative task.

	Swedish-German (N=46)		**Swedish-Turkish (N=48)**	
	MAIN1	MAIN2	MAIN1	MAIN2
≥1 sequence	82.6	82.6	62.5	64.6
no sequence	17.4	17.4	37.5	35.4

Next, GAOs were analyzed for the German and Turkish data. The omnibus test of model coefficient for the model of GAO versus other (including no sequence) was not significant ($\chi^2(3, N = 564) = 1.429$, $p = .699$). This means that the model including the variables language and age group did not explain a significant part of the variation in the data. There was thus no effect of language or age group on the production of GAOs in German and Turkish. The Swedish-German and the Swedish-Turkish bilinguals produced equal amounts of GAOs in their minority language narratives, and so did children in the different age groups. This lack of consistent age effects can be seen clearly in Table 7.23, which shows proportions of GAOs produced by the children in the different age groups by language and narrative task.

Table 7.23. German and Turkish sequences, percentage (%) GAOs out of all possible sequences, by language group and narrative task.

	Swedish-German (N=46)		Swedish-Turkish (N=48)	
	MAIN1	MAIN2	MAIN1	MAIN2
4-year-olds	7.1	4.8	2.1	4.2
5-year-olds	4.2	4.2	6.3	6.3
6-year-olds	6.3	10.4	2.1	8.3

In Table 7.24, an overview is given of children producing at least one GAO and those who produced no GAO for the two MAIN narratives. Only a small proportion of the children produced at least one GAO. There was no difference between the proportion of children in the two language groups producing at least one GAO, neither for MAIN1 ($\chi^2(1, N = 94) = .009$, $p = .926$), nor for MAIN2 ($\chi^2(1, N = 94) = 1.088$, $p = .297$).

Table 7.24. German and Turkish GAO-sequences, percentage (%) children producing at least one GAO-sequence, by language group and narrative task.

	Swedish-German (N=46)		Swedish-Turkish (N=48)	
	MAIN1	MAIN2	MAIN1	MAIN2
≥1 GAO	17.4	21.7	16.7	31.3
no GAO	82.6	78.3	83.3	68.8

7.3.4 Production of macrostructure: Summary

In Swedish, there were differences between all three age groups on macrostructure production scores both for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) – six-year-olds had higher scores than five-year-olds, who in their turn performed better than four-year-olds. The Swedish-Turkish bilinguals had lower macrostructure production scores than the monolinguals and Swedish-German bilinguals and fell around one year behind their monolingual and Swedish-German peers. This was mainly due to the Swedish-Turkish children producing attempts and outcomes to a lower extent than the children in the other two language groups.

Whereas the Swedish-Turkish bilinguals had equivalent scores in Swedish and Turkish, the Swedish-German bilinguals performed better in Swedish than in German for both MAIN1 and MAIN2. The Swedish-German bilinguals' German scores were higher than the Swedish-Turkish bilinguals' scores in Turkish for MAIN1, but not for MAIN2. In neither German nor Turkish were there any age effects on MAIN1 scores. In German, the six-year-olds performed better on MAIN2. On Turkish MAIN2, there was only an overall age effect, but no differences between specific age groups.

Attempts and outcomes were produced to a relatively large extent by all groups, and in all three languages. Other types of components were less fre-

quently produced, rarely exceeding 20% in any group. For both narrative tasks, most development was seen for attempts and outcomes, i.e. the components produced to quite a high extent already at age 4. Notably, there was no increase in the production of goals with age.

In Swedish, the Swedish-Turkish bilinguals produced a lower proportion of macrostructural sequences (sequences containing at least two of the core macrostructural components, goal, attempt or outcome) than the children in the other language groups. All three language groups produced similar proportions of goal-attempt-outcome-sequences. The German narratives also contained a higher proportion of sequences than the Turkish ones, but the proportion of children who produced at least one GAO did not differ between the two languages. In all three languages, there were clear age effects for proportions of sequences – the six-year-olds produced a higher proportion of sequences and the five-year-olds performed better than the four-year-olds. Interestingly, there was an increase with age in the use of GAOs in Swedish, but not in Turkish or German.

In each of the three languages, narratives elicited with MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) had equal complexity and received similar macrostructure production scores.

7.4 Comparing comprehension and production

In this section, results from comprehension and production of macrostructure are compared. The following questions were asked with regard to the comprehension and production scores:

- How does the children's production of macrostructural components compare to their comprehension?
- Is the relationship the same for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats)?

Additionally, there is a specific focus on the children's comprehension and production of goals.

7.4.1 Analysis

For each language, analyses were carried out on MAIN1 and MAIN2 separately, for the combined data from the three age groups, and for Swedish from all three language groups together. Comprehension and production scores in proportions out of the maximum score (relative scores) were compared using paired-samples Wilcoxon signed-rank tests. The comprehension and production relative scores were also correlated. Numbers of goals in comprehension and production were compared using paired-samples t-tests.

7.4.2 Swedish

Figure 7.29 shows the Swedish results (as percentage out of the maximum score) for comprehension and production of macrostructure for MAIN1 and MAIN2, by language group. Unsurprisingly, the related-samples Wilcoxon signed-rank tests showed that the children scored significantly better in comprehension than in production both for MAIN1 ($Z = 7$, $p < .001$) and MAIN2 ($Z = 179$, $p < .001$) in Swedish. Relative comprehension and production scores (in proportion out of maximum) for all children were significantly correlated for both MAIN1 ($r = .499$, $p < .001$) and MAIN2 ($r = .627$, $p < .001$), which indicates that children who scored better in Swedish comprehension tended to score better in Swedish production and vice versa.

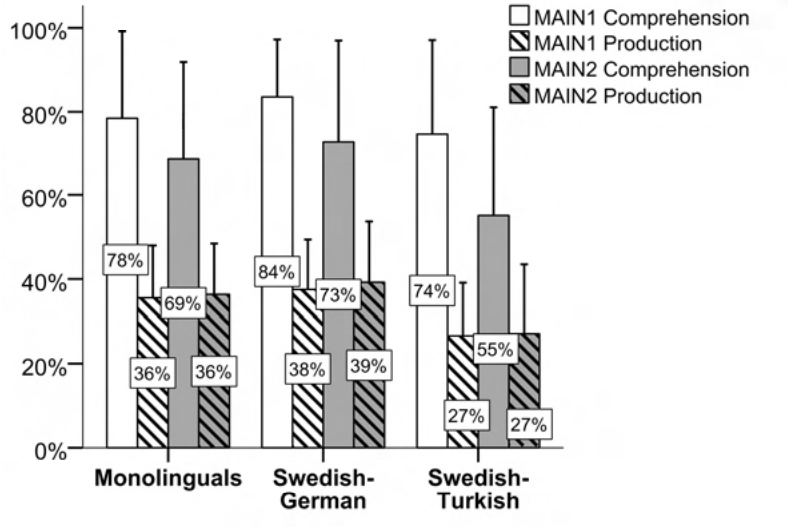

Figure 7.29. Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) comprehension and production of macrostructure, mean relative scores (in % out of the maximum score), by language group. Error bars show +1 SD.

Table 7.25 shows the mean number of goals in the Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) comprehension and production for the three language groups. Patterns are the same in each language group, with substantially higher mean number of goals for comprehension than for production in both narrative tasks. The paired-samples t-test showed that the children were significantly better at understanding than producing goals, both for MAIN1 ($t(162) = 27.406$, $p < .001$) and MAIN2 ($t(163) = 1.832$, $p < .001$). The results presented here thus show that mono- and bilingual Swedish-speaking children at this age are able to correctly comprehend (most of) the story characters' goals, but that they rarely produce them in their own narratives.

Table 7.25. Swedish narratives: Mean number of goals (SDs) produced in MAIN1 comprehension, MAIN1 production, MAIN2 comprehension and MAIN2 production, by language group.

	MAIN1 comp	MAIN1 prod	MAIN2 comp	MAIN2 prod
Monolinguals (N=72)	2.6 (0.6)	0.4 (0.7)	2.3 (0.7)	0.5 (0.6)
Swedish-German bilinguals (N=46)	2.7 (0.5)	0.8 (0.7)	2.5 (0.7)	0.6 (0.6)
Swedish-Turkish bilinguals (N=48)	2.7 (0.6)	0.7 (0.8)	1.9 (0.9)	0.6 (0.7)

Note. Comp = comprehension, prod = production. Max = 3. MAIN1 = Cat/Dog, MAIN2 = Baby Birds/Baby Goats.

7.4.3 German and Turkish

Figure 7.30 shows results (as percentages out of the maximum score) for comprehension and production of macrostructure for German and Turkish MAIN1 and MAIN2, by language.

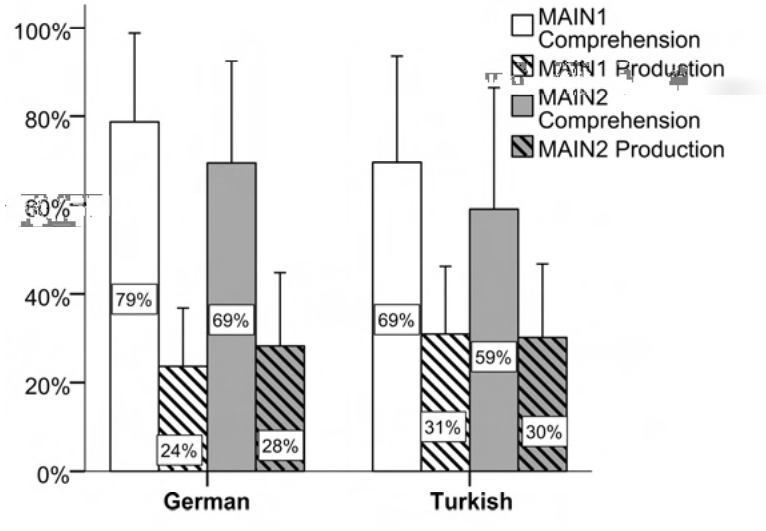

Figure 7.30. German/Turkish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) comprehension and production of macrostructure, mean relative scores (in % out of the maximum score), by language. Error bars show +1 SD.

Unsurprisingly, the related-samples Wilcoxon signed-rank tests showed that, in both German and Turkish, the children scored significantly better in comprehension than in production for MAIN1 (German: $Z = 1$, $p < .001$; Turkish: $Z = 55$, $p < .001$) and MAIN2 (German: $Z = 41$, $p < .001$; Turkish: $Z = 137.5$, $p < .001$). Relative comprehension and production scores (in proportions out of maximum) were not significantly correlated for any of the two

MAIN narratives, neither in German (MAIN1: r = -.024, p = .874; MAIN2: r = -.168, p = .265) nor in Turkish (MAIN1: r = -.035, p = .815; MAIN2: r = .029, p = .847). In the minority languages German and Turkish, contrary to what was found for Swedish, children who scored better in comprehension did not necessarily score better in production. A child could score high in minority language comprehension without scoring high in production and the other way around.

Table 7.26 shows the mean number of goals in MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) comprehension and production for German and Turkish. The pattern was the same for the two languages. The children understood a significantly higher number of goals than they spontaneously produced in their narratives, both in MAIN1 (German: $t(45) = 12.137$, $p < .001$; Turkish: $t(46) = 11.421$, $p < .001$) and in MAIN2 (German: $t(45) = 13.583$, $p < .001$; Turkish: $t(46) = 7.640$, $p < .001$). The pattern found for German and Turkish was thus the same as for Swedish.

Table 7.26. German and Turkish narratives: Mean number of goals (SDs) produced in MAIN1 comprehension, MAIN1 production, MAIN2 comprehension and MAIN2 production, by language.

	MAIN1 comp	**MAIN1 prod**	**MAIN2 comp**	**MAIN2 prod**
German (N=46)	2.5 (0.7)	0.6 (0.8)	2.5 (0.7)	0.7 (0.7)
Turkish (N=48)	2.3 (0.8)	0.5 (0.7)	1.9 (1.0)	0.5 (0.7)

Note. Comp = comprehension, prod = production. Max = 3. MAIN1 = Cat/Dog, MAIN2 = Baby Birds/Baby Goats.

7.4.4 Summary of results

In all three languages studied, Swedish, German and Turkish, children scored significantly better in comprehension than in production, when relative scores were compared. The pattern was the same for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) and the difference between comprehension and production was large. In Swedish, scores on comprehension and production were significantly correlated, whereas this was not the case for German and Turkish. This indicates that the relationship between comprehension and production may not be the same in majority and minority languages. Results also showed that, in all three languages and for both narrative tasks, the children understood a significantly higher number of goals than they produced in their narratives.

7.5 Discussion

In this chapter, the following six research questions were asked: (1) How does *comprehension* of narrative macrostructure develop from age 4 to 6?

(2) How does *production* of narrative macrostructure develop from age 4 to 6? (3) Do comprehension and production of narrative structure in Swedish differ between the three language groups? (4) Do the bilinguals perform differently in their two languages (Swedish vs German; Swedish vs Turkish)? (5) Are there any differences in narrative macrostructure between the two minority languages German and Turkish? (6) Are there differences between comprehension and production and between the two narrative tasks (MAIN1 Cat/Dog vs MAIN2 Baby Birds/Baby Goats)? In this section, the results are discussed. The discussion is divided into five different topics: development with age (Section 7.5.1), differences between the tasks (Section 7.5.2), differences between the language groups in Swedish (Section 7.5.3), the bilinguals' performance in their two languages (Section 7.5.4), and the performance in the two minority languages German and Turkish (Section 7.5.5). In the final section, some concluding remarks are made (Section 7.5.6).

7.5.1 Development with age

The first issue to be discussed is age development. Although there was a general trend for older children scoring higher than younger children, results varied between tasks and languages. There were stronger age effects in the majority language Swedish than in the minority languages German and Turkish. For Swedish comprehension, age effects were also stronger for MAIN2 (Baby Birds/Baby Goats) than for MAIN1 (Cat/Dog), whereas age effects in production were the same for both Swedish narratives tasks: six-year-olds performed significantly better than five-year-olds, who in turn had significantly higher scores than four-year-olds. The age effects found for Swedish are in line with the findings from other studies using MAIN (e.g. Bohnacker, 2016), as well as studies using other types of stimulus materials (e.g. Berman & Slobin, 1994; Mäkinen, 2014; Schneider et al., 2006; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Trabasso et al., 1992). For example, in both languages of Swedish-English bilinguals, Bohnacker (2016) found differences between age 5 and age 6–7. Macrostructure production scores in the current study were also similar to those of earlier MAIN-studies (cf. Table 7.1).

In both Turkish and German, there were no significant differences between the age groups on MAIN1 (Cat/Dog) comprehension or production, and age effects on MAIN2 (Baby Birds/Baby Goats) were less pronounced than in Swedish. For example, in production, the six-year-olds performed better on MAIN2 in German, and in Turkish, there was only a significant overall effect of age group, but no differences between any two age groups.

What could be the reason for the less clear age effects in the minority languages? A likely explanation is related to the children's general language proficiency, and especially the fact that children with relatively limited mi-

nority language proficiency can be found in all age groups.[153] It may well be the case that some of the older children have too limited minority language proficiency and therefore cannot show their full narrative competence on these tasks. This argument is strengthened when the children's performance in the two languages is compared (cf. e.g. Tables 7.12 and 7.13 for macrostructure production scores in German/Swedish and Turkish/Swedish). For example, a number of five- and six-year-old Swedish-German bilinguals performed well on Swedish macrostructure, but poorly in German.

In the current study, development with age was not analyzed for the individual comprehension questions. In a recent study investigating performance on different types of MAIN comprehension questions in the 72 monolinguals of the current study and the 52 Swedish-English bilinguals from Bohnacker (2016), Bohnacker & Lindgren (submitted) found that development with age is not identical for the different questions. Not only is the total comprehension score influenced by the child's age, but which questions a child can be expected to answer correctly varies with age; and such development does not neccessarily show in the total comprehension score (cf. Bohnacker & Lindgren, submitted). This issue should be investigated more in detail in future studies.

In line with earlier studies (e.g. Bohnacker, 2016; Kapalková et al., 2016; Soodla & Kikas, 2010; Trabasso et al., 1992), the children in the current study often included attempts and outcomes in their narratives, whereas other components, such as goals, settings and internal states, were realized less frequently, rarely exceeding 20% (out of a possible 100% for each component) in any group. Internal states as reaction were especially infrequent. Production of the different components varied somewhat between the two narrative tasks, which is likely linked to properties of the stimulus materials. However, the general trends were the same for both tasks. Interestingly, age development was also most clearly seen in attempts and outcomes, and there was a notable lack of increase in goals. These patterns were the same in all three languages. It thus seems that from age 4 to 6, children mainly develop their ability to include all the visible actions that form the core of the narratives. Development in the components that need to be inferred from the pictures, such as goals and internal states, is less pronounced. This is in line with the results from the study by Bohnacker (2016), who found that even children aged 6–7 produced few goals when telling the Baby Birds/Baby Goats story in Swedish and English (cf. also Soodla & Kikas, 2010), and with those of Trabasso et al. (1992), who found for English-speaking monolinguals that not until age 9 goals were produced as often as by adults.

Interestingly, in all three languages, there were clear age effects for overt realizations of sequences of at least two of the core macrostructural components (goals, attempts, outcomes) – the six-year-olds produced a higher pro-

[153] That this is indeed the case has been shown for vocabulary above (Chapter 5).

portion of sequences than the younger children and the five-year-olds performed better than the four-year-olds. The four-year-olds often included only single components and not sequences. This means that the older children produced narratives that had a higher level of macrostructural complexity than the narratives of the younger children. Thus, although there were few clear differences between the age groups for the macrostructure production scores in German and Turkish, the older children were in fact better at producing sequences of macrostructural components than the younger children. Some aspects of macrostructure do seem to develop from age 4–6 also in the minority languages, but a total score for production of different types of macrostructural components does not fully catch this development.

Irrespective of age and language group, when the children produced sequences of the central story components, these were mainly AO-sequences (so-called 'action sequences') and not, for example, full episodes (GAO-sequences). GAOs were relatively rare in all age groups in all three languages. With the exception of the Swedish-German six-year-olds, who produced 18.8% GAOs in their Swedish MAIN1 narratives, only around 5–10% of the episodes were realized as GAOs, and percentages were even lower for the four-year-olds. These figures are comparable to those of Bohnacker (2016), who reported 7% and 12% for Swedish-English bilinguals aged 5 and 6–7, respectively. The current study found an age development in the use of GAOs in Swedish, but not in Turkish or German. It should be pointed out that although there was a development with age for the use of GAOs in Swedish, and the increase in the proportions of GAOs was relatively steep (cf. Trabasso & Nickels, 1992), percentages of GAOs remained low at age 6. Thus, even the narratives produced by the six-year-olds in the current study are far from adultlike.

7.5.2 Task effects

Before considering the differences between the two narratives tasks, MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), the results for comprehension versus production should be discussed. Irrespective of language (Swedish, German, or Turkish), the children had higher relative scores in comprehension and understood a significantly higher number of goals than they produced in their narratives. The pattern was the same for both narrative tasks and the difference between comprehension and production was large. Generally, the children performed well on comprehension, although certain comprehension questions were harder, as shown by lower accuracy (cf. Section 7.2.3). For the oldest children, scores on the comprehension questions were approaching the maximum possible score, whereas no group had mean scores on production that were higher than 50% of the maximum. These results are in line with findings from earlier studies showing that narrative comprehension develops earlier than production (e.g. Bohnacker, 2016;

Roch et al., 2016; Stein & Glenn, 1979; Trabasso et al., 1992). The children's performance varied somewhat between the different comprehension questions. Regarding goal comprehension, even the youngest children, the four-year-olds, showed good comprehension of goals, while the same children extremely rarely included goals in their narratives. This is in line with findings from earlier studies (Bohnacker, 2016; Stein & Glenn, 1979; Trabasso et al., 1992).

Interestingly, the relationship between comprehension and production scores was not the same in the three studied languages. In the majority language Swedish, comprehension and production were significantly correlated for both narrative tasks, e.g. scoring higher on comprehension meant scoring higher on production as well. In the two minority languages, German and Turkish, there was no correlation between scores on comprehension and production. This indicates that the relationship between comprehension and production may be different for the majority and the minority languages. What could be the reason for this minority-majority language discrepancy? If the ability to understand and express narrative macrostructure is completely language-independent, one would expect the pattern to be the same in bilinguals' both languages. However, answering the comprehension questions is a less demanding task linguistically and cognitively, and it could be the case that children with a lower level of language proficiency in the minority language can still score relatively well on comprehension.

The second aspect of task effects concerns differences between MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats). In the studies that have used Cat/Dog for retelling and Baby Birds/Baby Goats for telling (Kunnari et al., 2016; Maviş et al., 2016; Roch et al., 2016), scores from Cat/Dog and Baby Birds/Baby Goats were compared and differences between the children's scores on the two narrative tasks attributed to the elicitation mode (retelling vs telling). The current study has found differences in comprehension scores even when Cat/Dog and Baby Birds/Baby Goats are both used in the same (telling) mode. In all three languages (Swedish, German, Turkish), the children scored better on Cat/Dog than on Baby Birds/Baby Goats in comprehension, whereas there were no such task effects in production. Why then is Cat/Dog comprehension, but not production, better than Baby Birds/Baby Goats? A number of possible explanations spring to mind.

First, it could be the case that the comprehension questions of Cat/Dog are easier than those of Baby Birds/Baby Goats, whereas the actual story content has a similar level of difficulty. Although the questions were constructed to be parallel, there could still be differences in wording that influence how difficult they are. Second, the order of the tasks might be an issue. The order of the narrative tasks was not counterbalanced in the current study; Cat/Dog was always told first and Baby Birds/Baby Goats later. However, it remains unclear why this would make scores on the comprehension questions higher on Cat/Dog but not affect the production scores.

In addition to the differences between MAIN1 and MAIN2, there was also a difference in comprehension between Baby Birds and Baby Goats, but only for Swedish. Performance on the comprehension questions was better for Baby Goats, except for in the Swedish-Turkish group. Baby Birds and Baby Goats may thus not be completely equivalent stories in comprehension. No such difference was found for MAIN1; the children performed virtually identically on Cat and Dog.

7.5.3 Effects of language group in Swedish

In Swedish comprehension, all three language groups (monolinguals, Swedish-German bilinguals and Swedish-Turkish bilinguals) scored equally well on MAIN1 (Cat/Dog), but the Swedish-Turkish bilinguals performed lower than the other two language groups on MAIN2, but only on the Baby Goats story; for Baby Birds performance of all three groups was the same. This indicates that groups may differ from each other on some types of stimulus material, but not on others.

In Swedish production, there were clear differences between the Swedish-Turkish bilinguals, who performed lower, and the Swedish-German bilinguals and the monolinguals, who performed similarly. In terms of overall scores, the Swedish-Turkish bilinguals were about one year behind their monolingual and Swedish-German peers, with e.g. the Swedish-Turkish six-year-olds performing lower than the monolingual five-year-olds. Interestingly, the Swedish-Turkish children produced goals and internal states, i.e. components thought to be more difficult, to the same extent as the monolinguals and the Swedish-German bilinguals. However, they failed to consistently produce attempts and outcomes, i.e. the 'simple' components or the 'action base' of the episodes, which led to lower overall scores. The Swedish-Turkish bilinguals' narratives also contained fewer macrostructural sequences (sequences containing at least two of the core macrostructural components, goals, attempts or outcomes). Swedish-Turkish children reached the highest level of performance, as measured by at least one GAO-sequence as often as Swedish monolinguals and Swedish-German bilinguals. A relatively large proportion of the Swedish-Turkish bilinguals produced no sequence of any kind (i.e. the lowest level of complexity). It is possible that these differences between the Swedish-Turkish bilinguals and the other two groups are due to linguistic difficulties (i.e. in lexicon or grammar).

Why might the Swedish-Turkish bilinguals have performed lower in Swedish than the children in the other language groups? A number of possible reasons come to mind. These are more or less the same reasons as discussed in the chapters on vocabulary (Chapter 5) and character introduction (Chapter 6). For example, overall lower language proficiency in Swedish may prevent the Swedish-Turkish children from expressing story content or answering the comprehension questions in a sufficiently clear manner in

order to score a point. The link between vocabulary production, an aspect of general language proficiency, and macrostructure is analyzed in the following chapter.

7.5.4 The bilinguals' performance in their two languages

The next point to be discussed concerns the bilinguals' performance in their two languages. In comprehension, both bilingual groups performed similarly in their two languages. Results were, however, different for production. Whereas the Swedish-Turkish bilinguals had equivalent scores in Swedish and Turkish, in line with earlier studies showing similar performance in the two languages (e.g. Bohnacker, 2016; Fiestas & Peña, 2004; Kunnari et al., 2016; Pearson, 2002), in production, the Swedish-German bilinguals performed better in Swedish than in German on both narrative tasks. For some older Swedish-German bilinguals, who had relatively limited proficiency in German (as indicated by vocabulary production scores, cf. Chapter 5), the difference between production scores in the two languages was large. This indicates that not all bilingual groups perform similarly on macrostructure in their two languages, and that age is not the only factor affecting child performance. The results also suggest that practice may somewhat compensate for limited language proficiency, but only in comprehension; there was an order effect in German comprehension and a reasonable explanation for this effect is practice. In narrative production, whether or not German was tested first or second did not influence the children's production scores. A reasonable explanation is that children with relatively weak German are able to utilize their knowledge from the first testing in the simpler task of answering the comprehension questions, but not in the more complex storytelling task. Interestingly, no such effect was found for the Swedish-Turkish bilinguals' Swedish comprehension scores. This indicates that practice may not affect all groups of children similarly, just as was the case in the study by Bohnacker (2016), who found a practice effect for 6–7-year-old Swedish-English bilinguals, but not for 5-year-olds.

7.5.5 Performance in the minority languages

The final research question concerns the comparison of the bilinguals' performance in their respective minority language, German and Turkish. In comprehension, for both narrative tasks, the Swedish-German bilinguals performed better in German than the Swedish-Turkish bilinguals did in Turkish. In MAIN1 (Cat/Dog), both groups comprehended goals equally well. In production, the Swedish-German bilinguals' German scores were higher than the Swedish-Turkish bilinguals' Turkish scores on MAIN1, but not on MAIN2 (Baby Birds/Baby Goats). A closer look at the types of components and at macrostructural complexity revealed that the German narra-

tives contained more attempts and outcomes as well as more sequences of the core macrostructural components (goals, attempts, outcomes), but that there was no difference between the two languages in the proportion of children who produced at least one GAO, similar to the results for Swedish.

There are different explanations for the Swedish-German children's good performance in German as compared to the Swedish-Turkish bilinguals in Turkish. First, the Swedish-German children may be helped in telling their narratives and in answering the comprehension questions in an understandable manner by the fact that Swedish and German are closely related languages that share many cognates. Using Swedish words in the German testing may in some cases lead to answers or narrative productions that are comprehensible to a German speaker with no knowledge of Swedish and in such cases the child may score a point on macrostructure. Many Swedish-German children may thus be able to score relatively well on the comprehension questions, and to some extent also on narrative production, despite having a relatively limited expressive vocabulary in German. The same goes for the child's ability to understand the question asked as well – a better understanding of what is asked naturally makes it easier to answer the question correctly. The Swedish-Turkish children do not have the same type of benefit, since their two languages have little in common. In order to bring more clarity to this issue, a detailed investigation of both correct and incorrect answers to the comprehension questions would be needed.

Second, it remains an open question how similarities between languages in terms of specific linguistic and narrative structures influence the children's performance on narrative tasks. Earlier studies have shown there are differences in how narrative events are expressed across languages (cf. Berman & Slobin, 1994c; Strömqvist & Verhoeven, 2004), but to my knowledge, no study has investigated how these differences may influence bilinguals. Such a study would not only need to study typical constructions used in narratives and how they differ between languages but also compare bilingual children with different language combinations but matched on e.g. overall level of language proficiency.

Third, it may also be the case that telling a story and correctly answering comprehension questions, which is taken here as a proxy for narrative comprehension, is a general ability that is more dependent on other aspects than a high level of language proficiency. Two such potential aspects are familiarity with storytelling activities and the ability to adapt to the experimental setting. However, it remains unclear how the two studied bilingual groups differ in this regard.

7.5.6 Concluding remarks

To conclude, the children generally had high scores on narrative comprehension, but scored much lower on narrative production. This indicates that

narrative production should not be used as a measure of how well children understand narrative content, as is sometimes done, e.g. in the Bus Story Test (Renfrew, 1969), but that it is necessary to use probe questions to assess the children's comprehension in a more direct manner. This is especially true for goals which were rarely produced, but which were understood well by most children. Additionally, there were differences between the two narrative tasks (MAIN Cat/Dog and Baby Birds/Baby Goats) in comprehension, but not in production. This interesting finding should be investigated further, especially since MAIN Cat/Dog and Baby Birds/Baby Goats were constructed to be parallel (cf. Gagarina et al., 2012, pp. 20–25).

Effects of age for macrostructure were smaller in the minority languages German and Turkish than in the majority language Swedish. Results for macrostructure production and to some extent also comprehension point to differences between the two bilingual groups, differences that are especially pronounced in the majority language Swedish. Just as for vocabulary (Chapter 5) and character introduction (Chapter 6), the Swedish-German bilinguals performed similarly to the monolinguals on both comprehension and production of the two narrative tasks (MAIN1 Cat/Dog, MAIN2 Baby Birds/Baby Goats), whereas the Swedish-Turkish bilinguals only performed comparably to the other two language groups on the MAIN1 total comprehension score and on MAIN2 goal comprehension. The Swedish-German bilinguals performed better in German than the Swedish-Turkish bilinguals for comprehension and MAIN1 production; scores for MAIN2 production did not differ between the two minority languages. Interestingly, the main reason for the Swedish-Turkish bilinguals' lower total scores in production of macrostructure was their lower production of attempts and outcomes. This shows the importance of analyzing different types of macrostructural components and not only total scores (cf. Bohnacker, 2016).

8 Linking vocabulary and character introduction to macrostructure

As described above (Chapter 1), narrative competence is a complex set of abilities, involving different aspects of language as well as more general cognitive abilities. Three different aspects of narrative ability, vocabulary, character introduction and narrative macrostructure have been reported on in the current study (Chapters 5–7). The question remains whether there is a relationship between these aspects. In this final results chapter, the links between macrostructure production scores and character introduction on the one hand, and on the other, between macrostructure production scores vocabulary, are investigated. Investigating these links constitutes a first step in analyzing which factors influence production of narrative macrostructure.

8.1 Vocabulary and macrostructure

In this section, effects of vocabulary production (CLT scores) and narrative vocabulary (number of different words, NDW) on macrostructure scores are analyzed for MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) in all three languages, Swedish, German and Turkish. In all analyses, age of the child was also included. The following research question was asked:

- Do vocabulary production and narrative vocabulary influence macrostructure production scores when the age of the child is controlled for?

8.1.1 Analysis

To analyze the effect of vocabulary (NDW and CLT scores) and age on macrostructure production scores, linear regression models were carried out for the two narrative tasks separately in each language, with MAIN macrostructure production score as the dependent variable, and age in months, NDW, and vocabulary production score (CLT) as independent variables (predictors). For each of the models, it was also tested whether the length of the narrative (total number of words, TNW) influenced the macrostructure production score. For Swedish, data from all three language groups, the monolinguals, the Swedish-German bilinguals and the Swedish-Turkish

bilinguals, was combined, as the focus was on the effects of vocabulary and age on the macrostructure production scores irrespective of mono- or bilingual status.

8.1.2 Swedish

Table 8.1 shows the final linear regression model for the Swedish MAIN1 (Cat/Dog) macrostructure production score, which includes age in months, number of different words (NDW) and vocabulary production score (CLT) as predictors. All of these have a highly significant effect on the MAIN1 macrostructure production score (*p*s < .001). Children who were older or produced a higher number of different words or had a higher score on the CLT, also scored higher on macrostructure production. The model explains 54.1% of the variance in the children's scores. Adding TNW as predictor did not significantly improve the model ($\Delta r^2 < .001$, $p = .807$). This shows that the length of the narrative does not have an impact on the macrostructure score, whereas the number of types, a measure of lexical diversity does influence the score significantly.

Table 8.1. Summary of linear regression model 8.1: Swedish MAIN1 (Cat/Dog) macrostructure production score, all children (N=165).

Predictor	***β***	***SE***	***t***	***p* value**
Intercept/Constant	-4.721	.814	-5.802	< .001***
Age (months)	.076	.013	5.944	< .001***
NDW	.043	.011	4.019	< .001***
Vocabulary production score (CLT)	.083	.013	6.157	< .001***
Model evaluation	$R^2 = .541$			

Note. *** = p < .001. NDW = number of different words.

The final linear regression model for the Swedish MAIN2 (Baby Birds/Baby Goats) macrostructure production score is shown in Table 8.2. The results for MAIN2 are similar to those for MAIN1, with the model explaining 53.9% of the variance in scores and with all three factors being significant.[154] In Swedish, the variables age in months, NDW, and CLT score thus all influence the macrostructure production scores regardless of whether the child tells MAIN1 (Cat/Dog) or MAIN2 (Baby Birds/Baby Goats). Narrative vocabulary (measured by number of different words) and vocabulary production (measured by the score on the test CLT) thus affect the children's score on narrative macrostructure independently, and in addition to the child's age. Although the two measures of vocabulary are significantly correlated with each other and with age (recall Chapter 5), all three factors thus independent-

[154] Similarly to the results for the MAIN1, adding TNW as a predictor to the MAIN2 model did not significantly improve model fit ($\Delta r^2 = .009$, $p = .082$).

ly contribute to explaining the child's macrostructure production score in Swedish.

Table 8.2. Summary of linear regression model 8.2: Swedish MAIN2 (Baby Birds/Baby Goats) production score, all children (N=166).

Predictor	***β***	***SE***	***t***	***p* value**
Intercept/Constant	-4.702	.924	-5.091	< .001***
Age (months)	.036	.015	2.375	.019*
NDW	.114	.015	7.758	< .001***
Vocabulary production score (CLT)	.081	.013	6.475	< .001***
Model evaluation	R^2 = .539			

Note. * = p < .05, *** = p < .001. NDW = number of different words.

8.1.3 German

In Table 8.3, the summary of the linear regression model for the German MAIN1 (Cat/Dog) macrostructure production scores is shown. The predictor age was not significant, in accordance with the results when the effect of age on the MAIN1 macrostructure scores was measured in isolation (see Section 7.3.2.2). Both vocabulary production scores and NDW were significant: children with higher German CLT scores and those producing narratives with higher NDW had higher German macrostructure production scores.[155] The model explains almost 70% of the variance in scores, showing that NDW and CLT scores are two very important factors for the German MAIN1 macrostructure scores.

Table 8.3. Summary of linear regression model 8.3: German MAIN1 (Cat/Dog) macrostructure production score, Swedish-German bilinguals (N=46).

Predictor	***β***	***SE***	***t***	***p* value**
Intercept/Constant	-1.920	1.403	-1.369	.178
Age (months)	.011	.020	.516	.609
NDW	.091	.017	5.240	< .001***
Vocabulary production score (CLT)	.060	.027	2.251	.030*
Model evaluation	R^2 = .696			

Note. *** = p < .001, * = p < .05. NDW = number of different words.

Table 8.4 shows the final linear regression model for the German MAIN2 (Baby Birds/Baby Goats) macrostructure production scores. The model explains 56.8% of the variation in the children's scores. Interestingly, effects were not the same as for the German MAIN1. Age did significantly influence MAIN2 scores, in line with results presented in Section 7.3.2.2. There was no effect of NDW, but CLT scores had a highly significant effect. This

[155] Adding TNW to the model did not significantly improve model fit (Δr^2 < .009, p = .278).

means that for MAIN2, the number of different words a child produced did not influence the macrostructure production scores, whereas older children and children who had high German CLT scores scored higher on macrostructure.[156]

Table 8.4. Summary of linear regression model 8.4: German MAIN2 (Baby Birds/Baby Goats) macrostructure production score, Swedish-German bilinguals (N=46).

Predictor	***β***	***SE***	***t***	***p* value**
Intercept/Constant	-4.807	1.751	-2.745	.009**
Age (months)	.056	.026	2.113	.041*
NDW	.035	.025	1.361	.181
Vocabulary production score (CLT)	.127	.030	4.317	< .001***
Model evaluation	R^2 = .568			

Note. * = p < .05, ** = p < .01, *** = p < .001. NDW = number of different words.

8.1.4 Turkish

The final linear regression model for the Turkish MAIN1 (Cat/Dog) macrostructure production scores is shown in Table 8.5.

Table 8.5. Summary of linear regression model 8.5: Turkish MAIN1 (Cat/Dog) macrostructure production score, Swedish-Turkish bilinguals (N=48).

Predictor	***β***	***SE***	***t***	***p* value**
Intercept/Constant	-3.342	2.008	-1.665	.103
Age (months)	.063	.029	2.162	.036*
NDW	.032	.033	.959	.343
Vocabulary production score (CLT)	.052	.028	1.902	.064
Model evaluation	R^2 = .280			

Note. * = p < .05. NDW = number of different words.

The model in Table 8.5 explained only 28% of the variance in the children scores, and the only significant predictor was age.[157] This was strikingly different from the results for Swedish and German. There thus appear to be some other factors than the two vocabulary measures investigated here that are of importance for how well the Swedish-Turkish bilinguals perform on MAIN1 macrostructure.

In Table 8.6, the summary of the linear regression model for the Turkish MAIN2 (Baby Birds/Baby Goats) macrostructure production scores is shown. Contrary to the results for Turkish MAIN1, all predictors had a significant effect on the macrostructure production scores for MAIN2: higher CLT scores, higher NDW and the child being older all led to higher MAIN2

[156] Adding TNW to the model did not significantly improve model fit (Δr^2 = .005, p = .488).
[157] Adding TNW to model did not improve model fit (Δr^2 < .001, p = .897).

scores.[158] The model explained 54.4% of the variance in the Swedish-Turkish children's Turkish MAIN2 scores.

Table 8.6. Summary of linear regression model 8.6: Turkish MAIN2 (Baby Birds/Baby Goats) macrostructure production score, Swedish-Turkish bilinguals (N=48).

Predictor	***β***	***SE***	***t***	***p* value**
Intercept/Constant	-6.028	2.037	-2.959	.005**
Age (months)	.066	.030	2.198	.033*
NDW	.058	.020	2.870	.006**
Vocabulary production score (CLT)	.104	.027	3.892	< .001***
Model evaluation	R^2 = .544			

Note. * = p < .05, ** = p < .01, *** = p < .001. NDW = number of different words.

8.2 Character introduction and macrostructure

Introducing characters properly is important for telling a narrative that is understandable to the listener. In this section, the relationship between the ability to introduce story characters and the macrostructure production score is investigated. The following research question was asked:

- Is there a difference in the Swedish MAIN1 macrostructure score between children who introduce characters appropriately and those who do not?

The analyses were carried out on the Swedish (Section 8.2.1) and German (Section 8.2.2) on data from the MAIN1 (Cat/Dog) only, since character introduction has not been analyzed for MAIN2 and for Turkish.

8.2.1 Analysis

Data from all children (monolinguals and the two bilingual groups for Swedish, all three age groups for both Swedish and German) were combined. Age group and language group were not included as variables in the analysis. For each language (Swedish and German) separately, two one-way ANOVAs were run with MAIN1 macrostructure production scores as the dependent variable: the first had the number of characters introduced in a fully appropriate manner (i.e. using indefinite NPs or equivalent expressions, see Section 6.3.1) as the independent (grouping) variable, and the second had number of lexical NPs as the independent (grouping) variable. The reason for performing the second analysis was that lexical NPs, irrespective of

[158] Adding TNW did not improve model fit (Δr^2 = .031, p = .085).

their morphological form, make it clear to the listener which the story characters were, and it was thought that this may be as important as using the appropriate morphological form. As described above (cf. Section 3.2, Table 3.2), MAIN1 (Cat/Dog) has three animate characters and every child (monolinguals, Swedish-German bilinguals, Swedish-Turkish bilinguals) told one MAIN1 narrative in Swedish. Similarly, each Swedish-German child told one MAIN1 narrative in German. This means that, in each language, each child had three opportunities to introduce a character.

8.2.2 Swedish

Table 8.7 shows the Swedish MAIN1 (Cat/Dog) macrostructure scores where the children have been grouped by the number of fully appropriate referring expressions (FAE) they used to introduce story characters. It was relatively uncommon to not use any fully appropriate expression at all (only 27 children out of 165). Around one third of the children exclusively used fully appropriate expressions for character introduction.

Table 8.7. Swedish MAIN1 (Cat/Dog) macrostructure score and character introduction, by the number of characters introduced with fully appropriate referring expressions (FAE), all children (N=165).

	0 FAE (N = 27)	**1 FAE (N = 36)**	**2 FAE (N = 43)**	**3 FAE (N = 59)**
Mean (SD)	3.5 (1.9)	5.2 (2.1)	5.9 (2.4)	6.7 (1.7)
Range	0 – 9	1 – 9	0 – 11	3 – 11

Note. N = number of children. Max score = 17.

There was a significant effect of the number of fully appropriate character introductions on the macrostructure score ($F(3, 161) = 16.366$, $p < .001$). The subsequent post-hoc tests showed that children who did not use any fully appropriate NPs to introduce story characters performed significantly worse than those who used one or more fully appropriate expressions. Additionally, those who used only one fully appropriate expression performed lower than those who used only fully appropriate expressions. Children using two fully appropriate expressions did not differ from those who used one or three fully appropriate expressions.[159]

Table 8.8 shows the Swedish MAIN1 (Cat/Dog) macrostructure scores by the number of lexical NPs used to introduce story characters. All children used at least one lexical NP to introduce the story characters, and it was relatively uncommon to use only one lexical NP (13 children). A large majority of the children (118 children) exclusively used lexical NPs. As can be seen from the score ranges, low-scoring children were found in all groups, indicating that just because a child uses lexical NPs to introduce story characters,

[159] P-values for all pairwise comparisons are found in Appendix 6, Table A6.1.

he/she does not necessarily tell a story that includes many macrostructural components. In all three groups, there was substantial variation in scores, but the children who scored highest all used at least two lexical NPs.

Table 8.8. Swedish MAIN1 (Cat/Dog) macrostructure score and character introduction, by the number of characters introduced with a lexical NP (lexNP), all children (N=165).

	1 lexNP (N = 13)	**2 lexNP (N = 34)**	**3 lexNP (N = 118)**
Mean (SD)	3.6 (2.2)	4.5 (2.3)	6.2 (2.0)
Range	0 – 7	0 – 10	0 – 11

Note. N = number of children. Max score = 17.

The number of lexical NPs used was significantly linked to the child's macrostructure score ($F(2, 162) = 14.770$, $p < .001$). The post-hoc tests showed that there was only a significant difference between the children who used only lexical NPs and those who did not; when a child used only lexical NPs, that child tended to score higher. Whether or not a child used one or two lexical NPs did not make a difference for the macrostructure score.[160]

8.2.3 German

In Table 8.9, the German MAIN1 (Cat/Dog) macrostructure scores are shown for children using different numbers of fully appropriate NPs.

Table 8.9. German MAIN1 (Cat/Dog) macrostructure score and character introduction, by the number of story characters introduced with fully appropriate referring expressions (FAE), Swedish-German bilinguals (N=46).

	0 FAE (N = 6)	**1 FAE (N = 10)**	**2 FAE (N = 14)**	**3 FAE (N = 16)**
Mean (SD)	3.3 (1.9)	4.1 (1.8)	5.7 (2.6)	6.4 (2.5)
Range	0 – 5	1 – 7	0 – 9	3 – 10

Note. N = number of children. Max score = 17.

There was a significant difference in the mean macrostructure score between the groups ($F(3, 42) = 3.744$, $p = .018$). The post-hoc tests showed that there was only a difference between children who used only fully appropriate expressions and children who did not; the former performed significantly better than the others.[161]

Table 8.10 shows the results for the number of lexical NPs in the German MAIN1 (Cat/Dog). Since the number of children using only one lexical NP was very small (N=4) and the means for this group and the children using two lexical NPs were similar, the statistical analysis compared the children

[160] For p-values of all pairwise comparisons, see Appendix 6, Table A6.2.
[161] For p-values for all pairwise comparisons, see Appendix 6, Table A6.1.

using only lexical NPs to those who did not. The independent-samples t-test showed a significant difference between the groups ($t(44) = -2.288$, $p = .027$): children who introduced all characters in their German MAIN1-narrative with lexical NPs scored better than those who did not.

Table 8.10. German MAIN1 (Cat/Dog) macrostructure score and character introduction, by the number of characters introduced with a lexical NP (lexNP), Swedish-German bilinguals (N=46).

	1 lexNP (N = 4)	**2 lexNP (N = 10)**	**3 lexNP (N = 32)**
Mean (SD)	3.8 (1.0)	4.2 (2.5)	5.8 (2.5)
Range	3 – 5	0 – 8	0 – 10

Note. N = number of children. Max score = 17.

8.3 Summary of results and discussion

In this chapter, the relationship between vocabulary and macrostructure and character introduction and macrostructure has been investigated. In Swedish and German,[162] children who used more lexical NPs and more fully appropriate expressions to introduce story characters also scored higher on the MAIN1 (Cat/Dog) macrostructure production.[163] Not surprisingly, children who were able to introduce all story characters with fully appropriate expressions were also better at including macrostructural elements in their narratives. Both these abilities were connected to age, as has been shown in Chapter 6 for character introduction and in Chapter 7 for macrostructure; older children were better at introducing story characters and at telling story content in an understandable manner. In future studies, age should be included as covariate in analyses of the relationship between the ability to introduce story characters and the ability to express narrative macrostructure.

However, age is not the only component influencing the macrostructure score. This became clear when effects of vocabulary on the macrostructure score were analyzed. Age was a significant predictor for the children's scores on both MAIN-tasks in Swedish and Turkish, and on German MAIN2. The two vocabulary measures, NDW and CLT scores also contributed significantly to explaining variance in the macrostructure scores (with the exception of both factors for Turkish MAIN1 and NDW for German MAIN2). With the exception of Turkish MAIN1, where only age was significant, models including these three factors explained at least 54% of the variance in the children's macrostructure production scores. Although there are

[162] Character introduction was not analyzed in the Turkish data.
[163] Note that a lenient score for reference was used (see Section 7.3.1); it was not the case that introducing a character with e.g. a pronoun automatically led to a lower macrostructure score.

clearly other factors contributing to the children's macrostructure scores, age and vocabulary have thus been shown to be important ones.

The effects of age, NDW and CLT scores were slightly different for the majority language Swedish and the two minority languages German and Turkish. Results for German and Turkish were not the same either. All three factors were significant in Swedish, but for the minority languages, this was only the case for Turkish MAIN2. Group size and large individual variation within the age groups may to some extent explain why not all of these factors were significant for all narratives in the minority languages. Other factors may influence narrative macrostructure in the minority languages more than vocabulary does, for example morphology and/or syntax.

It is not surprising that the number of different words a child uses in a narrative significantly may affect the macrostructure score – with a very low number of different words it is not possible to clearly describe the different actions in the MAIN pictures. The link between vocabulary in narratives and story structure score has been shown in a few earlier studies (e.g. Heilmann et al., 2010; Uccelli & Páez, 2007). What is interesting is that general vocabulary knowledge as measured by the production score on the CLT affects the scores in addition to NDW in the narrative. This supports the conclusion from Chapter 5 that these two types of measures assess different types of vocabulary knowledge. There are some possible reasons for why a child with a better general knowledge of words may score higher on narrative macrostructure. First, it could be the case that having a higher CLT score is linked to a better command of the language generally. Children with high CLT scores may have higher language proficiency in general, which may enable them to tell narratives that are easier to understand and where macrostructural components are expressed more clearly. Second, children with high CLT scores may be better at using words more accurately, i.e. choosing more appropriate words to describe a specific action. How well-chosen the specific words are is not reflected in the number of different words, although using a higher number of different words may mean that more specific words are used. How words are used to convey specific actions should be investigated in future studies of children's narrative vocabulary.

Interestingly, narrative length (as measured by the total number of word tokens, TNW) did not explain a significant part of the variance in the macrostructure score for any of the tasks in any of the three languages, when added to the model with age, NDW and CLT scores. This means that telling a longer narrative does not necessarily lead to a higher macrostructure score, but rather what matters is whether a wide range of lexical items is used.

To conclude, children who introduced characters appropriately, older children, and children with high vocabulary production scores and who produced high numbers of different words in their narratives tended also to have higher macrostructure production scores.

9 Overview of results

In this chapter, concise overviews are given of the most important results for each of the four results chapters, vocabulary (Chapter 5), character introduction (Chapter 6), macrostructure (Chapter 7), and the link between vocabulary and character introduction and macrostructure (Chapter 8).

9.1 Vocabulary

In Chapter 5, vocabulary production scores (CLT) and narrative vocabulary measured by the number of different words (NDW) in the MAIN narratives were investigated in all three languages. Below, a summary of the most important findings is given.

Vocabulary production (CLT)

- In Swedish, the six-year-olds performed significantly better than the two younger groups.
- In German and Turkish, there was no clear effect of age group.
- The Swedish-Turkish bilinguals performed significantly lower than the children in the other two language groups.
- The Swedish-Turkish six-year-olds performed lower than the monolingual four-year-olds.
- The Swedish-German bilinguals performed significantly better in Swedish than in German, whereas the Swedish-Turkish bilinguals had higher scores in Turkish than in Swedish.
- There was no difference between the scores of the Swedish-German bilinguals in German and of the Swedish-Turkish children in Turkish.

Narrative vocabulary (NDW)

- In Swedish MAIN1 (Cat/Dog), there were only differences between the six-year-olds and both younger groups. In Swedish MAIN2 (Baby Birds/Baby Goats), there were differences between all three age groups
- In German, there was only an effect of age on MAIN2.
- There were no age effects in Turkish.
- There was no effect of language group on MAIN2 NDW in Swedish. On MAIN1, the Swedish-Turkish bilinguals had a lower NDW than the monolinguals, but there were no differences between the bilingual

groups or between the Swedish-German bilinguals and the monolinguals.
- In Swedish, NDW was significantly correlated with CLT scores for the monolinguals and Swedish-Turkish bilinguals, but not for the Swedish-German bilinguals. In both German and Turkish, there were strong positive correlations between NDW and CLT scores.

9.2 Character introduction

In Chapter 6, character introductions in MAIN1 Cat/Dog and ENNI A2/B2 were investigated in Swedish and German. The main findings were the following:

- In Swedish, the six-year-olds used significantly lower proportions of pronouns than the younger groups. For proportions of fully appropriate NPs, there were differences between all three age groups with the older children performing better than the younger ones.
- In German, the six-year-olds performed better than the younger children.
- In Swedish, the Swedish-Turkish children used significantly higher proportions of pronouns and lower proportions of fully appropriate NPs than the other groups. There was no difference between the Swedish-German bilinguals and the monolinguals.
- The Swedish-Turkish six-year-olds performed similarly to the monolingual and Swedish-German five-year-olds.
- There was no difference between the Swedish-German bilinguals' performance in German and Swedish.
- In both Swedish and German, the children performed significantly better on MAIN Cat/Dog than on ENNI A2/B2.

9.3 Narrative macrostructure

In Chapter 7, comprehension and production of narrative macrostructure as well as the relationship between comprehension and production were investigated. Data from MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) for Swedish, German and Turkish was analyzed. The main findings for each of these topics are listed below.

Comprehension
- In Swedish, differences were found between the six-year-olds and the younger groups on MAIN1, and between all three age groups on MAIN2.

- In Turkish and German, there was no difference between the age groups on MAIN1, and for MAIN2, there were only differences between the four- and the six-year-olds in Turkish, and an overall effect of age group in German.
- In Swedish, the language groups performed similarly on MAIN1, but the Swedish-Turkish bilinguals performed lower on MAIN2.
- There were no differences between the two languages for any of the bilingual groups.
- The Swedish-German bilinguals performed higher in German than the Swedish-Turkish bilinguals did in Turkish.
- In all three languages, performance was higher on MAIN1 than on MAIN2.

Production

- In Swedish, there were clear differences between all three age groups on the overall scores, with older children performing better.
- In Turkish and German, there was no effect of age group on MAIN1 overall scores. On MAIN2, there was only a difference between the six-year-olds and both younger groups in German and an overall effect of age group in Turkish.
- In Swedish, the Swedish-Turkish bilinguals had lower overall scores and produced attempts and outcomes to a lower extent than the children in the other two language groups; the Swedish-Turkish six-year-olds performed at the same level as the monolingual five-year-olds.
- In terms of overall scores, the Swedish-German bilinguals performed better in Swedish than in German. The Swedish-Turkish bilinguals performed similarly in both languages.
- The Swedish-German bilinguals performed better in German than the Swedish-Turkish bilinguals did in Turkish on MAIN1, but not on MAIN2.
- Most age development was seen on attempts and outcomes.
- There was no age development in the inclusion of goals.
- There was a clear development with age in the use of sequences of macrostructural components in all three languages.
- GAOs (complete episodes) were relatively rare in all age groups in all three languages, but in Swedish there was a development with age in the use of GAOs.
- The Swedish-German bilinguals produced more sequences of macrostructural components in German than the Swedish-Turkish bilinguals did in Turkish, but both groups produced GAOs to the same extent in their respective minority languages.
- There were no differences in performance between MAIN1 and MAIN2.

Comparison of comprehension and production

- In all three languages and for both MAIN1 and MAIN2, there was a much higher performance in comprehension than in production.
- The children rarely produced goals, but understood them well.
- In Swedish, but not in German and Turkish, there were significant correlations between scores on comprehension and production.

9.4 Linking vocabulary and character introduction to macrostructure

In Chapter 8, the links between vocabulary and macrostructure on the one hand and character introduction and macrostructure on the other were investigated. The results are listed below.

Vocabulary and macrostructure

- Age, vocabulary production scores (CLT) and number of different words (NDW) explained significant portions of the variance in the children's macrostructure production scores for both MAIN1 (54%) and MAIN2 (54%) in Swedish, and for Turkish MAIN2 (54%).
- For Turkish MAIN1, 28% of the variance in scores could be explained by age.
- For German MAIN1, vocabulary production scores and NDW explained 70% of the variation in macrostructure production scores.
- For German MAIN2, age and vocabulary production scores explained 57% of the variance in macrostructure production scores.
- Narrative length in total number of words (TNW) was not a significant predictor when included in models together with total number of different words (NDW), age and vocabulary production scores (CLT).

Character introduction and macrostructure (only investigated for MAIN Cat/Dog)

- In both Swedish and German, children who used more lexical NPs and fully appropriate expressions to introduce the story characters scored higher on the MAIN (Cat/Dog) macrostructure production.

10 General discussion and conclusions

This chapter contains a general discussion of the results as well as conclusions. The following four research questions were asked.

- **RQ1:** How and when do the different aspects of narrative ability (vocabulary, character introduction, macrostructure) develop between age 4 and 6?
- **RQ2:** To what extent is this development similar for mono- and bilingual children and for different bilingual groups?
- **RQ3:** Are there differences between the bilingual children's two languages?
- **RQ4:** Are there differences in the children's performance depending on the narrative task?

This chapter is divided into eight sections. In Section 10.1, development with age (RQ1) is discussed. Sections 10.2 and 10.3 deal with the performance in Swedish (RQ2) and the bilinguals' performance in their two languages (RQ3), respectively. Section 10.4 homes in on the bilinguals' performance in the two minority languages Turkish and German (RQ2). Section 10.5 revisits the relationship between different narrative aspects. Section 10.6 discusses issues related to task effects (RQ4) and general methodological reflections. In Section 10.7, limitations of the study are described and proposals for future studies are put forward. Finally, Section 10.8 concludes the thesis.

10.1 Development with age

The results for development of narrative macrostructure in the current study were similar to those of earlier studies using the MAIN Baby Birds/Baby Goats narrative task (e.g. Bohnacker, 2016; Gagarina, 2016; Kunnari et al., 2016; Maviş et al., 2016; Roch et al., 2016, see also Table 7.1), as well as studies using other stimulus materials (e.g. Berman & Slobin, 1994b; Mäkinen, 2014; Schneider et al., 2006; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Trabasso et al., 1992). The ability to introduce characters appropriately in narrative discourse seems to develop relatively early (cross-linguistically) in Swedish-speaking children, for both mono- and bilinguals,

compared to results from earlier studies of other languages (e.g. Hickmann et al., 1996; Kail & Hickmann, 1992; Kail & Sanchez y Lopez, 1997; Küntay, 2002), something that may also be related to the specific methodology used in the different studies.

Importantly, the current study contributes to the research into age effects by showing that development with age was different in the majority language, Swedish, compared to the two minority languages, German and Turkish. In the majority language Swedish, there was a clear development with age for all the studied aspects. The six-year-olds performed better than the younger children on all measures in Swedish. They had higher vocabulary production scores (on the CLT), produced a higher number of different words in their MAIN narratives, used fewer pronouns and more fully appropriate NPs to introduce story characters in MAIN1 (Cat/Dog) and ENNI (A2/B2), and scored higher both in comprehension and production of macrostructure for MAIN1 and MAIN2 (Baby Birds/Baby Goats) than both younger groups. On a number of measures in Swedish, there were also significant differences between four- and five-year-olds (though not for proportions of pronouns used to introduce story characters, number of different words in MAIN1, scores on vocabulary production and MAIN1 comprehension). The results seem to indicate a clearer development from age 5 to 6 than from age 4 to 5. The variation in performance between individual children also tended to be larger in the younger groups with higher standard deviations and wider score ranges. Individual differences between six-year-olds were less pronounced.

In the minority languages, German and Turkish, development with age was much less clear. In fact, most measures showed no difference between the three age groups in German and Turkish, although in some cases performance did increase when age was analyzed as a continuous variable (i.e. the child's age in months). This is in line with findings from studies of vocabulary (Bohnacker et al., 2016; Dijkstra et al., 2016; Gagarina et al., 2014; Gathercole et al., 2013; Hoff et al., 2012; Uccelli & Páez, 2007). However, it partially contradicts studies of narrative macrostructure that showed clear differences between age groups in the minority language (e.g. Bohnacker, 2016 for Swedish-English bilinguals aged 5 vs 6–7; Roch et al., 2016 for Italian-English bilinguals aged 5–6 vs 6–7). Although the reasons behind the differences between studies remain unclear, they are likely linked to input patterns, including whether the children receive schooling in the minority language. In the minority languages, substantial individual variation in performance was found within all age groups. Similarly, the amount of input received also varied between individual children in the same age group. The results of the current study indicate the need for further studies investigating the relationship between input patterns, age, general language proficiency and performance on various narrative measures. Such studies should not

only investigate effects at group level, but also look more closely at profiles of individual children (cf. the case studies in Bohnacker et al., 2016).

An additional comment about the results for production of macrostructure and age is warranted. In Swedish, most age development was seen for attempts and outcomes and notably, there was no age development in the inclusion of goals. From age 4 to 6, children thus mainly seem to develop their ability to include the visible actions that form the core of the narratives, such as attempts and outcomes (cf. Berman & Slobin, 1994b, p. 62). Development in the components that need to be inferred from the pictures, such as goals and internal states, is less pronounced (cf. Bohnacker, 2016; Trabasso et al., 1992). There was a clear development with age in the use of sequences of macrostructural components in all three languages. Six-year-olds have a more developed ability to include the core aspects of the narrative, in the form of macrostructural sequences, than four-year-olds do. However, at age 6, full episodic structures were still not used very often. GAO-sequences, i.e. complete episodes, were relatively rare in all age groups in all three languages (cf. Bohnacker, 2016), but in Swedish there was a development with age in the use of GAOs (cf. Trabasso & Nickels, 1992).

To summarize the results of the current study with respect to age development, the aspects studied do not develop in an identical manner throughout the age range studied. Also, the development with age in the minority languages German and Turkish is not as clear as in the majority language Swedish, which is possibly linked to variation in the amount of input between individual children within the same age group. Although some aspects of macrostructure do develop from age 4 to 6 in the minority languages, a total score for production of different types of macrostructural components does not fully show this development. This is especially true when differences between age groups are analyzed. The study has shown that the late preschool years, i.e. age 4–6, are central for the development of the ability to introduce referents appropriately, with four-year-olds using some appropriate expressions, and (Swedish-German and monolingual) six-year-olds using such expressions almost exclusively. Comprehension of narrative macrostructure is already at a high level at age 4, but still develops substantially from age 4 to 6. Production of narrative macrostructure is very rudimentary at age 4, and even at age 6, it is still not at a very high level, with low proportions of complete episodic structures. Even the narratives produced by the six-year-olds in the current study are far from adultlike. Certain narrative components, such as goals and internal states, are rarely included in narratives by children at these ages, although these components are understood when explicitly probed.

To conclude, although important steps in the development of narrative competence are taken within the studied age range, at age 6, development is still far from complete.

10.2 The bilinguals' performance in Swedish

Two things are striking when looking at the children's performance in Swedish. First, the Swedish-German bilinguals performed like Swedish monolinguals on all the measures studied. Second, for all measures except narrative vocabulary and comprehension of macrostructure in MAIN1 Cat/Dog, the Swedish-Turkish bilinguals performed lower than the Swedish monolinguals. The difference between the groups was largest for vocabulary production, where, on average, the Swedish-Turkish bilinguals were more than two years behind their monolingual peers; the Swedish-Turkish six-year-olds performed lower than the monolingual four-year-olds. This large difference for vocabulary production is especially interesting in light of the lack of significant differences for narrative vocabulary. Despite knowing fewer words, as indicated by the score on the vocabulary task, the Swedish-Turkish bilinguals produced numbers of different words in their narratives that were comparable to the children in the other groups. For the other measures, character introduction, narrative comprehension in MAIN2 and production of narrative macrostructure, differences were smaller than for vocabulary production, but still substantial. For these measures, the Swedish-Turkish bilinguals were around one year behind their monolingual and Swedish-German peers at age 6. However, it is important to note that some Swedish-Turkish children *did* perform similarly to the children in the other groups. For most measures, individual variation was somewhat higher in the Swedish-Turkish group, with some children performing at a high level compared to their age peers in the other two groups and others performing very low.

For the production of narrative macrostructure, it should be pointed out that the Swedish-Turkish bilinguals did not perform lower than the other children in terms of goals or GAO-sequences. Instead, they produced fewer attempts and outcomes as well as fewer sequences of the core macrostructural components. This suggests that the narrative competence of the highest-performing Swedish-Turkish children is at the same level as for the children in the other groups, but that as a group, the Swedish-Turkish bilinguals were not able to express all depicted actions of the stories to the same extent. Future studies could investigate whether this is related to their lower lexical knowledge, as indicated by their lower score on vocabulary production (CLT).

There are a number of reasons why the Swedish-Turkish children, but not the Swedish-German children, performed lower than monolinguals in Swedish. First, it could be that a lower amount of Swedish *language input* leads to lower language proficiency in Swedish. The amount of input in a language has been shown to be an important factor for vocabulary development (e.g. Bohnacker et al., 2016; Gatt et al., 2017; Haman, Wodniecka, et al., 2017; Leseman, 2000; Potgieter & Southwood, 2016). The Swedish-Turkish children did receive a substantial amount of daily Swedish input according to

parental report, but the exact amount as well as the input quality remains unclear. In future studies, more detailed information about the input should be collected. In the current study, vocabulary production scores were clearly linked to narrative macrostructure, indicating that narrative ability is being influenced by general language proficiency. Overall lower proficiency in Swedish may prevent the Swedish-Turkish children from expressing story content or answering comprehension questions in an understandable manner.

A second explanation for the lower performance of the Swedish-Turkish is that the *language combination* influences acquisition, both of narrative structures and of language more generally. One area where the influence of the language combination seems clear is reference, more specifically character introduction (cf. Chen & Lei, 2013; Finnstedt, 2013; Jia & Paradis, 2015; Serratrice, 2007). Since the Swedish and German referential systems are similar whereas the Turkish is not, learning to introduce characters appropriately in Swedish might be more difficult for Swedish-Turkish than for Swedish-German bilinguals, due to interference between the two systems. Whether this also holds for other narrative aspects, such as the verbalization of macrostructural components, remains to be shown.

It is likely that input, and language combination together play a role in creating a context for the Swedish-German bilinguals that is boosting their performance in Swedish (and also German, cf. Section 10.4), and another context for the Swedish-Turkish bilinguals, one that is not as supportive to the acquisition of Swedish or narrative ability more generally.

To conclude, the results from the current study clearly show that not all bilingual groups perform the same and that it is therefore important not to put all bilinguals 'in one pot' and contrast them with monolinguals.

10.3 The bilinguals' performance in their two languages

Results for the bilinguals' performance in their two languages are somewhat mixed. The Swedish-German bilinguals had significantly higher vocabulary production scores in Swedish than in German, whereas the Swedish-Turkish bilinguals performed better on vocabulary production in Turkish than in Swedish. Both bilingual groups performed similarly in their two languages on narrative comprehension and the Swedish-Turkish group also on production of macrostructure. On narrative production, the Swedish-German bilinguals performed better in Swedish than in German, whereas they performed similarly in both languages for character introduction. The results of the current study differ somewhat from earlier studies that showed that bilinguals performed similarly in both their languages for narrative macrostructure (cf. e.g. Bohnacker, 2016; Fiestas & Peña, 2004; Kunnari et al., 2016; Pearson, 2002), but are in line with others showing higher scores in the language which the child has received more exposure to (better performance in

L1 Slovak than in L2 English; Kapalková et al., 2016; better performance in L1 Italian than in L2 English for 5-6-year-olds; Roch et al., 2016). The present study thus suggest that narrative macrostructure *is* related to general language proficiency (in the tested language) and not only to the development of general cognitive abilities, e.g. Theory of Mind or narrative schemata.

10.4 Comparing performance in the minority languages Turkish and German

For the performance of the bilingual groups in their respective minority language, German or Turkish, results are also mixed. Surprisingly, on no task did the Swedish-Turkish bilinguals perform higher in Turkish than the Swedish-German bilinguals did in German, even though the former group were reported to receive more parental input in their minority language and had higher parental ratings for their minority language proficiency than the latter. The bilingual groups performed equally well on vocabulary production and on production of macrostructure in MAIN2 (Baby Birds/Baby Goats) in their minority languages. The Swedish-German bilinguals performed higher in German than the Swedish-Turkish bilinguals did in Turkish for narrative comprehension on both MAIN tasks and also in production for MAIN1 (Cat/Dog). The Swedish-German bilinguals also produced more sequences of macrostructural components in German than the Swedish-Turkish bilinguals did in Turkish, although both groups produced GAOs to the same extent in their respective minority languages.

A number of possible explanations exist for the lack of differences between the groups on some aspects and for the Swedish-German bilinguals' better performance on other aspects. In fact, some of the same factors that may account for the differences between the bilingual groups in the majority language can also be used to explain the results in the minority language, namely SES (cf. e.g. Akoğlu & Yağmur, 2016) and language combination (or language distance). The fact that the Swedish-German bilinguals acquire two languages that are closely related may compensate for the fact that these children, on average, receive less input in German than the Swedish-Turkish bilinguals do in Turkish. For example, the fact that Swedish and German are closely related languages sharing many cognates may help the Swedish-German children in telling their narratives, and in answering the comprehension questions in an understandable manner. Additionally, it may help them score high on the CLTs (cf. Lindgren & Bohnacker, submitted). The Swedish-Turkish children do not have the same type of benefit as the Swedish-German bilinguals since their two languages have relatively little in common. In order to fully determine this, additional studies that investigate the

effect of language distance on children's narrative competence need to be conducted.

10.5 The relationship between different narrative measures

The current study has shown that older children, children who introduce characters appropriately, and children with high vocabulary production scores and who produce high numbers of different words in their narratives, tend to also have higher macrostructure production scores. Importantly, telling a longer narrative does not necessarily lead to a higher macrostructure score, but what matters is whether the child uses a wide variety of different lexical items. Although age is important for performance in the majority language, as could be expected, it is thus not the only factor influencing the children's scores on narrative tasks. Older bilinguals cannot always utilize the same narrative competence in both languages, as doing so also depends on aspects of language proficiency, such as vocabulary and grammar.

Not surprisingly, children who are able to introduce all characters in a story with fully appropriate expressions are in general better at telling stories. Using fully appropriate expressions, such as indefinite nominal phrases, may reflect a more developed ability to take the listener's knowledge into account (cf. e.g. Ariel, 1990; Gundel et al., 1993). For example, being able to understand that the listener cannot see the pictures may lead the child to a higher degree of explicitness in his/her rendering of narrative events, which in turn leads to a higher score on narrative macrostructure.

It is also not surprising, but still an important finding, that the number of different words a child uses in a narrative significantly affects the macrostructure score (cf. Heilmann et al., 2010; Uccelli & Páez, 2007). What is interesting, though, is that vocabulary knowledge as measured by the production score on the vocabulary task CLT affects the macrostructure scores in addition to number of different words (in a model controlling for the child's age). This means that a more general measure of the child's vocabulary is also a good indicator of how well the child manages to tell the story. Although vocabulary task scores were correlated with number of different words (at least for the monolinguals and the Swedish-Turkish bilinguals in Swedish, and for both Turkish and German), both these measures thus help explain the variation in the children's narrative macrostructure scores.

How different aspects of narrative ability are linked to each other has rarely been studied and should be investigated more in-depth in future studies. The analyses carried out in the current study only constitute a first step towards understanding the interplay between different narrative aspects, with some interesting results.

10.6 Task effects and methodological reflections

In this section, the effect of the type of narrative task on the children's performance will be discussed first, followed by general methodological reflections. A number of differences between the tasks (comprehension vs production of macrostructure, MAIN Cat/Dog vs MAIN Baby Birds/Baby Goats and ENNI A2/B2, vocabulary production vs narrative vocabulary) have been found in the current study. Interestingly, most of these differences are highly consistent across language and groups. Such task effects show that the specific tasks employed to test children's narrative ability can have a substantial influence on children's performance.

First, the difference in the children's performance on comprehension and production of narrative macrostructure was striking. In all three languages and for both narrative tasks (MAIN Cat/Dog and Baby Birds/Baby Goats), there was a much higher performance in comprehension than in production. For example, the children rarely included the characters' goals in their narratives, but understood them well when probed. These results are in line with earlier studies (e.g. Bohnacker, 2016; Trabasso et al., 1992), but are reported here for the first time in a study of mono- and bilingual (Swedish-speaking) children. These results show that narrative production cannot be used as an indicator of children's comprehension of narrative macrostructure. As is true for language in general, children understand much more of narrative structure than they are able to (spontaneously) produce. The lack of a correlation between comprehension and production scores in German and Turkish indicates that this may be true to an even larger extent for bilinguals in their minority language.

Second, there was a difference between the two vocabulary measures, vocabulary production scores (CLT) and narrative vocabulary, in how they differentiated between the language groups in Swedish. Whereas the Swedish-Turkish bilinguals performed much lower than the two other language groups on the standardized vocabulary task, the differences were less clear for narrative vocabulary (number of different words). Additionally, both vocabulary measures independently explained part of the variance in children's macrostructure production scores, which suggests that they do not tap into the same type of lexical competence. Importantly, the measures yield different results with regards to effects of age and language group.

Third, there was a substantial effect of the stimulus material on the children's ability to introduce story characters appropriately. Across age groups and languages (Swedish, German), as well as language groups for Swedish, children performed significantly better on MAIN Cat/Dog than on ENNI A2/B2. The differences between the tasks in terms of inappropriate pronouns and of fully appropriate NPs were large. The current study thus extends the findings of Lindgren (2018) for Swedish monolinguals and shows that both mono- and bilingual Swedish-speaking children have greater difficulties

with introducing characters appropriately in ENNI A2/B2 than in MAIN Cat/Dog. The choice of stimuli affects children's performance in character introduction and thus has implications for conclusions drawn regarding the age at which appropriate referring expressions are used consistently to introduce story characters.

The fourth and final point on the topic of task effects concerns differences in the children's performance on MAIN1 Cat/Dog and MAIN2 Baby Birds/Baby Goats. Interestingly, in all three languages, there was a higher performance on MAIN1 comprehension of macrostructure, but in production, there were no differences between MAIN1 and MAIN2. It is important to point out again that this interesting finding should be investigated further in light of MAIN Cat/Dog and Baby Birds/Baby Goats containing the same macrostructural components and having been constructed in a similar fashion to enable the same type of assessment (cf. Gagarina et al., 2012, pp. 20–25). The results from the current study indicate that one cannot take for granted that comprehension scores from Cat/Dog and Baby Birds/Baby Goats can be compared. This may be due to small differences in their plotlines and the types of actions performed by the protagonists (cf. Gagarina et al., 2015, p. 255) and/or in their comprehension questions. Additionally, the results from Swedish also showed that the difficulty of the comprehension questions for Baby Birds and Baby Goats may not be completely equivalent. Despite the fact that the stories and the accompanying comprehension questions were constructed to be parallel, the children scored better on Baby Goats comprehension than on Baby Birds. To conclude, in addition to overall macrostructure in terms of episodic structure, the specific content and characters may also affect children's performance on narrative tasks. The specific effects of such differences should be investigated systematically in future studies.

All methodological decisions in a research project can be discussed. Here, I will focus on two larger topics, how the elicitation method may influence the data, and how the type of analysis impacts on the results.

In the present study, a standardized procedure was used to administer the narrative tasks. The procedure was designed to make the data as comparable as possible and gave the children minimal support in telling the stories. There was no shared visual attention between child and experimenter and no prompting besides minimal backchanneling and general questions such as 'what happened then?'. This was a conscious decision in order to get insights into narrative competence without adults scaffolding the child. It is clear, though, that such a procedure makes the task more difficult and may not always elicit the children's best performance. For example, four-year-olds were able to produce goals when explicitly probed (e.g. answering comprehension questions), although they generally did not include them in their narratives. Not allowing the experimenter to give more specific prompting thus leads to lower performance. This type of situation is also somewhat

artificial in that it is relatively far from storytelling contexts in everyday life, where adults (presumably) adapt their responses and the type of support they give to the needs of the individual child. The fact that the experimenter was not allowed to help the child with e.g. lexical items for the story characters also put children with limited vocabulary at a disadvantage. The effect of the child's level of general language proficiency is thus relatively strong when this type of procedure is used, which means that what is measured is not only narrative competence, but also language proficiency more broadly. Yet, if the procedure is allowed to be very flexible and to vary with the child's needs, this actually means that a slightly different procedure is used with each child and that the data then are not fully comparable.

Additionally, a monolingual context was created in order to test the bilingual children's ability in each of their languages. This situation may not reflect the bilinguals' everyday situations, as they may be used to telling stories using words from both languages. Creating a monolingual context may thus underestimate the child's general narrative ability. Although setting up a monolingual context is the only way to make sure that the child's ability in *both* languages is assessed (in a 'bilingual context', the child may simply choose to use his/her strongest language), it must be kept in mind that neither testing fully taps into the bilinguals' complete narrative ability. Which procedure to use is a choice; there would have been valid reasons for making different choices than the ones in the current study. However, how the data were elicited must be kept in mind when comparing results from different studies.

The results are not only influenced by how the data were collected but also by the type(s) of analysis carried out. To name an area central to the current study, there are various ways in which macrostructure can be analyzed, based on different theoretical models. In the present study, macrostructure was coded (i.e. scored) following the MAIN model (cf. Sections 1.1 and 7.3.1.1), and analyses was carried out on the total scores, types of components and macrostructural sequences, with a specific focus on complete episodes (goal-attempt-outcome sequences). The choice in the present study to use this specific model was based on the fact that it was developed together with the stimulus material used. It was thus the logical choice. The question to ask is how this choice influenced the result. First, the MAIN model, just as any model, awards points for specific components and these components include specific aspects. For example, in the MAIN setting component, character introductions are not included, although one could argue that they form an essential part of setting the scene for the narrative events. In the current study, this potential shortcoming of the MAIN model was addressed through the inclusion of a separate analysis of character introductions. If character introductions had been included in the setting, it is likely that the results for this component would have been higher, since, at least in Cat/Dog, most characters were introduced by the children using lexical noun phrases.

Second, while it is undoubtedly a strength that the MAIN model clearly separates internal states as initiating event from goals, as this separation enables a detailed analysis of different types of inferences the children make when narrating, the MAIN definition of a complete episode is stricter than for example that of Stein & Glenn's (1979) story grammar model. In MAIN, only the goal can be the first component in a full episode, whereas, in the other story grammar models, it can also be an internal state as initiating event (IST as IE, an internal response in the terminology of Stein & Glenn) or even an initiating event (cf. Section 1.1). In the MAIN model, no point is awarded for describing the initiating event, and not every episode depicted in the MAIN picture sequence contains a clear initiating event. However, if episodes where the child produced an IST as IE together with attempt and outcome were to count as complete episodes, this would influence the figures for complete episodes substantially. In the current study, children rarely produced both IST as IE and goal in a single episode (cf. Section 7.3.3), but a number of children did use sequences consisting of IST as IE, attempt and outcome. Using the Stein & Glenn (1979) definition of an episode instead of the MAIN one would thus have meant that the children in the current study 'produced' higher proportions of complete episodes. Future studies could reanalyze the MAIN data using another model of macrostructure and compare it to the results from the MAIN model to measure the magnitude of the difference. To conclude, one can argue for the validity of one model over another, and in some cases two different options may be equally valid. However, it is important to remember that the theoretical definitions that form the basis for scoring the data have clear and important consequences for the results.

10.7 Limitations and future studies

The current study has a number of limitations. In addition to those that restrict the generalizations that can be made, concerning the number of participants, narrative aspects and language combinations studied, there are two main limitations. These will be discussed here.

First, the current study does not allow effects of SES to be tested. As all Swedish-German children came from high-SES backgrounds, it was not possible to separate SES and language combination. It therefore remains an open question whether the Swedish-German children's good performance is a result of their high SES. When recruiting participants, it was not possible to match the Swedish-Turkish and Swedish-German bilinguals on SES; these populations are simply too different in this regard. In order to investigate the influence of SES, language input and language combination further, it would be necessary to study other groups, for example children speaking either only typologically close *or* only typologically distant languages, but that

vary in terms of SES and amount of input, or children with different linguistic backgrounds that are SES-matched. Additionally, a larger group of monolingual children from low-SES backgrounds should be included in future studies, for a better comparison with low-SES bilinguals.

Second, all children were assessed on the narrative tasks in the same order (MAIN Cat/Dog followed by MAIN Baby Birds/Baby Goats and finally ENNI A2/B2). As the order of the tasks may influence performance, this limits the comparability of the tasks and the generalizability of task effects. It is possible that the task effects found are partly due to the order of the tasks. To give an example, having already carried out one narrative task may make the second one easier for the child, but a child might also perform less well later in the testing due to fatigue. In order to give conclusive evidence of task effects, the present study should be followed up by a study in which the order of the tasks is counterbalanced.

The current study opens up a number of avenues for further studies, some of which have already been mentioned in the discussion above as well as in Chapters 5–8. A few deserve specific mention here. First, it remains an open question at which age the Swedish-Turkish children will catch up with their monolingual and Swedish-German peers. Also, both the monolingual and the bilingual children's narratives are far from adultlike at age 6. Therefore, future studies should extend this research to older children, in order to find the age at which most types of components are included by the majority of children. As it is largely unknown what adult narratives elicited with MAIN look like, it would also be important to collect and analyze data from adults. Adult data could function as a kind of 'gold standard', indicating the level of narrative competence the children will be expected to develop towards. Second, it could be the case that the MAIN material with its complex (i.e. dense) structure of three complete episodes depicted in only six pictures makes it difficult for the children to consistently include all types of story components. In order to investigate this issue further it would be necessary to test children with stimulus materials that vary with regard to episodic complexity. Additionally, future studies should investigate further language combinations, in order to pinpoint how cross-linguistic differences may influence children's narratives. Studies should also be carried out on other narrative aspects than vocabulary, character introduction and macrostructure, such as temporal and causal cohesion. For example, future studies could look at the children's use of connectives or similar devices to link different parts of the narratives. Finally, the data from the current study could be analyzed using a more qualitative approach. This could, for example, entail an analysis of specific events in the narratives and how these are realized verbally by the children. A more holistic perspective could also be included, for example by rating children's narratives according to perceived overall quality (cf. Newman & McGregor, 2006). Adding such a perspective would enable us to gain more insights into specific details of children's narrative competence.

10.8 Conclusions

The present study has investigated the development of oral narrative competence from age 4 to 6 in Swedish monolinguals and in both languages of Swedish-German and Swedish-Turkish bilinguals. It is the first study of its kind in a Swedish context. Picture-based fictional narratives were elicited with two narrative tasks, Cat/Dog and Baby Birds/Baby Goats, from the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015) and one task, A2/B2, from the Edmonton Narrative Norms Instrument (ENNI; Schneider et al., 2005), using different but comparable stimuli in the bilinguals' two languages. Three different narrative aspects were analyzed: vocabulary, character introduction and (comprehension and production of) narrative macrostructure. The research questions concerned effects of age (RQ1), and similarities and differences between the different groups (RQ2), between the bilinguals' two languages (RQ3) and between narrative tasks (RQ4).

Already at age 4, most children performed well on narrative comprehension, but this ability continues to develop throughout the age range studied. Production of narrative macrostructure develops later and is at a rudimentary level at age 4. Although the older children performed better than the younger ones, even at age 6 children's narratives contained few complete episodic structures. The children mainly included actions visible in the narrative stimuli and only rarely verbalized goals and other macrostructural components that required them to draw inferences from the picture. The ability to introduce story characters appropriately in narratives developed strongly from age 4 to 6, but the choice of stimulus material had a large effect on performance. Vocabulary showed most improvement from age 5 to 6.

In the majority language Swedish, there was a clear development from age 4 to age 6 on all aspects of narrative competence studied, but this was not the case for the minority languages German and Turkish. Whereas in Swedish, variation was larger in among the younger children, in German and Turkish, large variation was found in all age groups.

Differences were found between the two bilingual groups, differences that were especially pronounced in the majority language Swedish. The Swedish-German bilinguals performed similarly to the monolinguals, whereas on most measures, the Swedish-Turkish bilinguals performed lower than the other two groups. The size of the group differences varied across measures, and to some extent also across tasks. Possible explanations for these differences between the two bilingual groups, such as differences in SES, general language proficiency, amount of language input, and effects of the language combination have been discussed throughout this thesis. It should be stressed that future studies will need to investigate the effects of these factors in detail in order to find out which aspects have the largest effect on the development of narrative competence. The relationship between the bilinguals' per-

formances in the two languages depended on the type of ability assessed, but generally, the Swedish-German children performed better in Swedish than in German, whereas the Swedish-Turkish children performed at the same level in both languages or slightly higher in Turkish. The two bilingual groups thus showed different patterns in their minority languages, in addition to differences in performance in the majority language Swedish.

To conclude, this thesis contributes to a better understanding of children's narrative competence and general (bilingual) language development. Bilinguals' two languages need not develop in parallel, and results may depend on the tasks and specific measures used. Bilingual groups differ from each other, and it is therefore not meaningful to compare all bilinguals to all monolinguals.

Appendix

Appendix 1 Examples of narratives

A1.1 MAIN1 (Cat/Dog)

MoSwe6-02, Macrostructure production score = 11 points

@Begin
@Languages: swe
@Participants: EXP Josefin_Lindgren Experimenter, CHI MoSwe602 Target_Child
@ID: swe|change_me_later|EXP||female|||Experimenter||
@ID: swe|change_me_later|CHI|6;6.|female|||Target_Child||
@Date: 13-OCT-2014
@Comment: MAIN, Cat
@Transcriber: Josefin Lindgren

*EXP:	här börjar sagan, titta på dom här två bilderna, nu får du berätta sagan för mig .
%eng:	here the story starts, look at these two pictures, now you can tell me the story
*CHI:	(.) det var en gång en katt <som> [/] (.) som såg en fjäril på en buske .
%com:	(.) once upon a time, there was a cat <who> [/] (.) who saw a butter-fly on a bush
*EXP:	aaa .
*CHI:	och den skulle försöka ta den .
%eng:	and it should try to take it
*EXP:	mmm .
%com:	CHI unfolds pic3-4, EXP helps her holding the pics
*CHI:	(.) men den fastna(de) i busken och fick ont .
%eng:	(.) but it got stuck in the bush and got pain
*EXP:	aaa .
*CHI:	(.) och den blev törstig .
%eng:	(.) and it became thirsty
*EXP:	mmm .
%com:	CHI unfolds pic5-6
*CHI:	(.) <och &to> [//] nej, den blev hungrig och åt upp fiskarna .
%eng:	(.) <and &too> [//] no, it became hungry and ate up the fish
*EXP:	aaa det gjorde den (.) vad hände det mera på bilderna ?
%eng:	aa, it did that (.) what happened more in the pictures
*CHI:	(.) &hm (.) gubben kommer med fiskar .
%eng:	(.) &hm (.) the guy comes with fish
*EXP:	aaa .
*CHI:	men sen ser han katten (.) och blir rädd och orolig .

%eng: but then he sees the cat (.) and becomes afraid and worried
*EXP: aaa .
*CHI: (.) och så ramlade bollen ner i vattnet .
%eng: (.) and so the ball fell into the water
*EXP: jaa .
*CHI: (.) så försöker han fiska upp den med fiskespöet och det funka(de) .
%eng: (.) so he tries to fish it out with the fishing rod and it worked
*EXP: aaa, då fick han upp den, titta (.) hände det nåt mer ?
%eng: aaa, then he got it out, look (.) did anything else happen
*CHI: &mmm näej .
%eng: &mmm, no
*EXP: näej, det var hela, titta, jättebra, vilken fin saga .
%eng: no, it was all, look, very good, what a nice story
@End

BiGer5-12, Macrostructure production score = 6 points

@Languages: swe
@Participants: EXP Josefin_Lindgren Experimenter, CHI BiGer512 Target_Child
@ID: swe|change_me_later|EXP||female|||Experimenter||
@ID: swe|change_me_later|CHI|5;8.|female|||Target_Child||
@Date: 31-AUG-2015
@Comment: MAIN, Dog
@Transcriber: Josefin Lindgren
*EXP: så tittar du först på alla bilderna .
%eng: then you first look at all pictures
*CHI: det var en hund .
%eng: it was a dog
%com: CHI starts without waiting for instructions, all pics are unfolded
*EXP: aa .
%com: EXP nods
*CHI: och en mus o(ch) o(ch) [/] den var där vid ett träd .
%eng: and a mouse and and [/] it was there next to a tree
*EXP: aa .
*CHI: o(ch) hunden är nära ett hus (.) .
%eng: and the dog is close to a house (.)
*EXP: (.) och sen ?
%eng: (.) and then ?
*CHI: och sen så <sen så> [/] &j jagar hunden musen, <och där och där> [//] och då &s springer musen in i ett hål på trädet, och <o(ch) o(ch)> [/] tittar ut där, så så [/] hunden slår sig .
%eng: and then so <then so> [/] &ch the dog chases the mouse, <and there and there> [//] and then the mouse runs into a hole in the tree, and looks out there, so so [/] the dog hits himself
%com: CHI points to mouse in the tree in pic3 when saying 'och tittar ut där'
*EXP: ooj .
%eng: ooh

*CHI: o(ch) o(ch) [/] då kom en pojke med en ballong mitt när han springer där, men <men men> [/] när han slår sig så råkar han tappa ballongen .
%eng: and and [/] then came a boy with a balloon just when he runs there, but <but but> [/] when he hits himself then he happens to drop the balloon
*EXP: aa .
*CHI: och och [/] så flyger den upp på trädet och så <och så och så> [/] klättrar han (.) och så har han tillbaka (.) sin sin [/] ballong, men men [/] hunden äter honoms korv .
%com: and and [/] so it flies up in the tree and so <and so and so> [/] he climbs (.) and so he has back his his [/] balloon, but but [/] the dog eats him's sausages
*EXP: oh åt han korvarna ?
%eng: oh, he ate the sausages
*CHI: &mm .
%com: CHI looks up from pics at EXP
*EXP: nämen, var det hela sagan ?
%eng: oh, that was the whole story
*CHI: &m
*EXP: aa titta jättebra var den, vilken fin saga !
%eng: aa, look it was great, what a nice story
@End

BiTur4-02, Macrostructure production score = 3 points

@Begin
@Languages: swe
@Participants: EXP Josefin_Lindgren Experimenter, CHI BiTur4-02 Target_Child
@ID: swe|BiLITAS|EXP||female|||Experimenter||
@ID: swe|BiLITAS|CHI|4;11.|female|||Target_Child||
@Date: 29-APR-2015
@Comment: MAIN1, Dog
@Transcriber: Sibylle Dillström
*EXP: titta, här börjar sagan, titta på dom här två bilderna, nu får du berätta sagan för mig .
%eng: look, here the story begins, look at these two pictures, now you can tell me the story
%com: EXP unfolds pic 1-2 and holds them in front of CHI
*CHI: &eh hunden försöker bita råttan .
%eng: &uh the dog tries to bite the rat
%com: CHI points to pic 2
*EXP: (.) aa, och sen ?
%eng: (.) aa, and then ?
%com: EXP unfolds pic 3-4
*CHI: (.) &ehm &ehm (.) hans ballong är där i trädet .
%eng: (.) &uhm &uhm (.) his balloon is there in the tree
%com: CHI points to pic 3
*EXP: ah, och sen ?
%eng: ah, and then

*CHI: (..) han blir rädd .
%eng: (..) he becomes scared
*EXP: aa, han blev rädd .
%eng: aa, he became scared
*CHI: och sen råttan skrattade .
%eng: and then the rat laughed
*EXP: aa, råttan skrattade, och sen ?
%eng: aa, the rat laughed, and then
%com: EXP unfolds pic 5-6
*CHI: (.) hunden skulle äta korv .
%eng: (.) the dog should eat sausage
%com: CHI points to pic 5
*EXP: aa, och sen ?
%eng: aa, and then
*CHI: och sen tog pojken ballongen .
%eng: and then the boy took the balloon
%com: CHI points to pic 6
*EXP: aa, och sen ?
%eng: aa, and then
*CHI: (.) &ehm (..) .
*EXP: hände det nåt mer ?
%eng: did anything else happen
*CHI: nej .
%eng: no
*EXP: var det hela ?
%eng: was it all
*CHI: &ah .
*EXP: ja, titta, jättebra var den .
%eng: yes, look, it was great
@End

A1.2 MAIN2 (Baby Birds/Baby Goats)

BiGer6-07, Macrostructure production score = 11 points

@Languages: swe
@Participants: EXP Josefin_Lindgren Experimenter, CHI BiGer607 Target_Child
@ID: swe|change_me_later|EXP||female|||Experimenter||
@ID: swe|change_me_later|CHI|6;10.|female|||Target_Child||
@Date: 06-FEB-2015
@Comment: MAIN, Baby Goats
@Transcriber: Josefin Lindgren
*CHI: det var en gång några bockar +/.
%eng: once upon a time, there were some billy-goats
%com: CHI starts telling without waiting for the instructions, EXP folds back pics, CHI holds pics
*EXP: mm .
*CHI: +, som villde [: ville] [*] ut på ängen .
%eng: who wanted out on the meadow
*EXP: mm .

*CHI: <den ena &lill> [//] minsta bocken kunde inte simma och ropade mamma, (.) mamma puttade ut barnet, och räven såg lilla barnet och tänkte äta barnet .
%eng: <the one &litt> [//] smallest billy-goat could not swim and shouted mother, (.) mother pushed the child out, and the fox saw the little child, and intended to eat the child
%com: CHI has a very dramatic tone towards end of utterance
*EXP: mm .
%com: EXP helps CHI unfold pic 3-4
*CHI: barnet (.) räven &spring (.) sprang fram, och &lill bockbarnet sprang iväg, men räven tog barnet i benet .
%eng: the child (.) the fox &ru (.) ran forward, and &little the billy-goat child ran away, but the fox caught the child in the leg
%com: CHI continues with a dramatic voice
*EXP: mm (.) .
%com: EXP nods and helps CHI unfold pic 5-6 and helps holding pics
*CHI: (.) då kom en Vogel [@g] och bet räven i svansen (.) räven släppte barnet (.) och sprang iväg, fågeln jagade (.) räven (.) in i djupa mörka skogen .
%eng: (.) then a bird came and bit the fox in the tail (.) the fox let go of the child (.) and ran away, the bird chased (.) the fox (.) into the deep dark forest
%com: CHI has a clear German pronunciation of Vogel ‘bird’
*EXP: mm hände det nåt mer ?
%eng: mm, did anything else happen
*CHI: nej det var allt, slut .
%eng: no, it was all, the end
%com: CHI folds back pics
*EXP: allt, slut, vad bra .
%eng: all, the end, what good
@End

BiTur6-08, Macrostructure production score = 7 points

@Begin
@Languages: swe
@Participants: EXP Maria_Johansson Experimenter, CHI BiTur6-08 Target_Child
@ID: swe|BiLITAS|EXP||female|||Experimenter||
@ID: swe|BiLITAS|CHI|6;7.|female|||Target_Child||
@Date: 22-NOV-2015
@Comment: MAIN2, Baby Birds
@Transcriber: Maria Johansson
*EXP: så nu får du berätta för mig .
%eng: so now you can tell me
%com: CHI holds the pictures while telling the story
*CHI: mamma fågeln (.) sa till (.) doms barn (.) ni får inte gå ut härifrån +/.
%eng: mother bird (.) said to (.) them’s child (.) you must not go out of here
*EXP: &mhm .

*CHI: ,+, annars flyger jag i väg .
%eng: otherwise I will fly away
*EXP: &mhm .
*CHI: så sen flugde [: flög] mamman i väg .
%eng: so then the mother flowed [: flew] away
*EXP: &m .
*CHI: sen villde [: ville] katten äta upp fåglarna .
%eng: then the cat wanted to eat up the birds
*EXP: &mhm (.) oj .
%com: EXP helps CHI to unfold pic 3-4
*CHI: sen mamma fågel skulle hämta mask till dom, sen katten klättrade upp (.) till trädet (.) sen tog katten en fågel (..) och sen, hunden kollade på katten .
%eng: then mother bird should get worms for them, then the cat climbed up (.) to the tree (.) then the cat took a bird (..) and then, the dog looked at the cat
%com: CHI unfolds pic 5-6
*EXP: &mhm .
*CHI: sen hunden drog kattens svans +/.
%eng: then the dog pulled the cat's tail
*EXP: oj .
%eng: oh
*CHI: +, &å och ät [: åt] den .
%eng: &o and eat [: ate] it
*EXP: &m .
*CHI: sen gjorde hunden så att katten fick gå i väg .
%eng: then the dog made so that the cat got to go away
*EXP: &mhm .
*CHI: snipp, snapp, snupp, sagan är slut .
%eng: snipp, snapp, snupp, the story is finished
*EXP: sagan är slut .
%eng: the story is finished
@End

MoSwe4-09, Macrostructure production score = 4 points

@Begin
@Languages: swe
@Participants: EXP Josefin_Lindgren Experimenter, CHI MoSwe409 Target_Child
@ID: swe|change_me_later|EXP||female|||Experimenter||
@ID: swe|change_me_later|CHI|4;4.|female|||Target_Child||
@Date: 09-SEP-2014
@Comment: MAIN, Baby Birds
@Transcriber: Josefin Lindgren
*EXP: här börjar sagan, titta på dom här två bilderna, nu får du berätta .
%eng: here the story begins, look at these to pictures, now you can tell
*CHI: mamma flög, sen kom den tillbaks .
%eng: mother flew, then it came back
*EXP: aaa (8.6) .

%com: EXP helps CHI to unfold pic3-4
*CHI: (.) sen kom katten upp .
%eng: (.) then the cat came up
*EXP: aaa .
*CHI: (.) och hon rivde [: rev] [*] barnet .
%eng: (.) and she scratched the child
*EXP: aaa .
*CHI: (4.9) dom får inte riva, katter .
%eng: (4.9) they are not allowed to scratch, cats
*EXP: näej .
%eng: no
*CHI: min katt gör inte det .
%eng: my cat does not do that
%com: CHI shakes her head repeatedly
*EXP: näej, det är bra .
%eng: no, that is good
*CHI: för jag känner igen den .
%eng: because I recognise it
%com: CHI probably intends to say ‘känner den’ = knows it, but instead says ‘känna igen’ = recognize
*EXP: aaa, det är en snäll katt .
%eng: aa, it is a nice cat
%com: EXP unfolds pic5-6
*CHI: och sen kom hunden och villde [: ville] [*] jaga katten .
%eng: and then the dog came and wanted to chase the cat
*EXP: aaa .
*CHI: (.) så &huuw (.) &eeh (.) bet hunden katten i svansen .
%eng: (.) so (.) &uuh the dog bit the cat in the tail
%com: CHI breaths in deeply between her teeth 'huuw'
*EXP: aaa, det gjorde han .
%eng: aa, that he did
*CHI: den var slut .
%eng: it was finished
*EXP: aaa vad hände sen ?
%eng: aa, what happened then ?
%com: EXP does not seem to hear CHI telling that it's the end
*CHI: jagar dom .
%eng: chases they/them
%com: CHI almost puts picture sequence down on the table
*EXP: aaa, (.) ja (.) var det hela ?
%eng: aaa, (.) yes (.) was it all ?
%com: CHI nods as answer to the question
*EXP: aa, titta, jättebra, vilken fin saga
%eng: aa, look, great, what a nice story
@End

Appendix 2 Narrative data

Table A2.1. Number of narratives, Swedish, German, and Turkish, by age group and narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)	Total
Swedish				
Monolinguals (N=72)	**72**	**72**	**72**	**216**
4-year-olds	24 (12/12)	24 (12/12)	24 (12/12)	72
5-year-olds	24 (12/12)	24 (12/12)	24 (12/12)	72
6-year-olds	24 (12/12)	24 (12/12)	24 (12/12)	72
Swedish-German bilinguals (N=48)	**45**	**46**	**45**	**136**
4-year-olds	14 (7/7)	14 (6/8)	13 (8/5)	41
5-year-olds	15 (7/8)	16 (8/8)	16 (8/8)	47
6-year-olds	16 (8/8)	16 (8/8)	16 (8/8)	48
Swedish-Turkish bilinguals (N=48)	**48**	**48**	**44**	**140**
4-year-olds	16 (8/8)	16 (8/8)	13 (6/7)	45
5-year-olds	16 (8/8)	16 (8/8)	15 (6/9)	47
6-year-olds	16 (8/8)	16 (8/8)	16 (9/7)	48
German				
Swedish-German bilinguals (N=48)	**46**	**46**	**44**	**136**
4-year-olds	14 (7/7)	14 (8/6)	13 (6/7)	41
5-year-olds	16 (8/8)	16 (8/8)	16 (8/8)	48
6-year-olds	16 (8/8)	16 (8/8)	15 (7/8)	47
Turkish				
Swedish-Turkish bilinguals (N=48)	**48**	**48**		**96**
4-year-olds	16 (8/8)	16 (8/8)		32
5-year-olds	16 (8/8)	16 (8/8)		32
6-year-olds	16 (8/8)	16 (8/8)		32

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

Table A2.2. Total number of words, Swedish narratives, by language group, age group and narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)	Total
Monolinguals (N=72)	**5,676**	**4,877**	**5,822**	**16,375**
4-year-olds	1,735	1,448	1,587	4,770
5-year-olds	1,783	1,536	1,978	5,297
6-year-olds	2,158	1,893	2,257	6,308
Swedish-German bilinguals (N=46)	**3,432**	**3,167**	**3,380**	**9,979**
Swedish	3,420	3,152	3,372	9,944
German	12	15	8	35
4-year-olds	967	808	778	2,553
Swedish	960	796	773	2,530
German	7	12	5	23
5-year-olds	991	1,057	1,145	3,193
Swedish	991	1,056	1,143	3,190
German	0	1	2	3
6-year-olds	1,474	1,302	1,457	4,233
Swedish	1,469	1,300	1,456	4,226
German	5	2	1	7
Swedish-Turkish bilinguals (N=48)	**3,637**	**3,859**	**4,144**	**11,640**
4-year-olds	1,118	1,113	1,122	3,353
5-year-olds	1,319	1,430	1,533	4,282
6-year-olds	1,200	1,316	1,489	4,005

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

Table A2.3. Total number of words, German narratives, by age group and narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)	Total
Swedish-German bilinguals (N=46)	**4,582**	**3,514**	**4,386**	**12,482**
German	4,500	3,438	4,262	12,200
Swedish	82	76	124	282
4-year-olds	1,369	1,043	1,135	3,547
German	1,337	1,021	1,096	3,454
Swedish	32	22	39	93
5-year-olds	1,396	1,046	1,513	3,955
German	1,360	997	1,450	3,807
Swedish	36	49	63	148
6-year-olds	1,817	1,425	1,738	4,980
German	1,803	1,420	1,716	4,939
Swedish	14	5	22	41

Note. N = number of children, BB = Baby Birds, BG = Baby Goats.

Table A2.4. Total number of words, Turkish narratives, by age group and narrative task.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	Total
Swedish-Turkish bilinguals (N=46)	**2,965**	**3,287**	**6,252**
Turkish	2,919	3,236	6,155
Swedish	46	51	97
4-year-olds	931	1,010	1,941
Turkish	915	994	1,909
Swedish	16	16	32
5-year-olds	976	1,208	2,184
Turkish	954	1,176	2,130
Swedish	22	32	54
6-year-olds	1,058	1,069	2,127
Turkish	1,050	1,066	2,116
Swedish	8	3	11

Note. BB = Baby Birds, BG = Baby Goats. N = number of children.

Table A2.5. Swedish words per narrative, pairwise comparisons between the language groups.

	MAIN2 (BB/BG)	ENNI (A2/B2)
Monolinguals vs Swedish-German	< .001***	1.00
Monolinguals vs Swedish-Turkish	.026*	.131
Swedish-German vs Swedish-Turkish	< .001***	.026*

Note. BB = Baby Birds, BG = Baby Goats, *** = p < .001, * = p < .05.

Table A2.6. Swedish words per narrative, pairwise comparisons between the age groups.

	MAIN1 (Cat/Dog)	MAIN2 (BB/BG)	ENNI (A2/B2)
4yrs vs 5yrs	1.00	.346	.079
4yrs vs 6yrs	.013*	.01*	.002**
5yrs vs 6yrs	.087	.468	.613

Note. BB = Baby Birds, BG = Baby Goats, ** = p < .01, * = p < .05.

Appendix 3 Vocabulary

Table A3.1. Swedish vocabulary production scores (CLT), pairwise comparisons between the language groups.

	p-value
Monolinguals vs Swedish-German	.406
Monolinguals vs Swedish-Turkish	< .001***
Swedish-German vs Swedish-Turkish	< .001***

Note. *** = p < .001.

Table A3.2. Swedish vocabulary production scores (CLT), pairwise comparisons between the age groups.

	p-value
4yrs vs 5yrs	.122
4yrs vs 6yrs	< .001***
5yrs vs 6yrs	< .001***

Note. *** = p < .001.

Table A3.3. Swedish MAIN1 (Cat/Dog) Number of Different Words (NDW), pairwise comparisons between the language groups.

	p-value
Monolinguals vs Swedish-German	.697
Monolinguals vs Swedish-Turkish	.001**
Swedish-German vs Swedish-Turkish	.088

Note. ** = p < .01.

Table A3.4. Swedish MAIN1 (Cat/Dog), and Swedish and German MAIN2 (Baby Birds/Baby Goats) Number of Different Words (NDW), pairwise comparisons between the age groups.

	Swe MAIN1	**Swe MAIN2**	**Ger MAIN2**
4yrs vs 5yrs	.434	.044*	1.00
4yrs vs 6yrs	< .001***	< .001***	.125
5yrs vs 6yrs	.008**	.011*	.015*

Note. *** = p < .001, ** = p < .01, * = p < .05

Table A3.5. Number of Different Words (NDW), Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), by age and language group.

	MAIN1	**MAIN2**
Monolinguals (N=72)		
Mean (SD)	46.1 (12.6)	40.8 (10.5)
Range	24 – 81	20 – 77
4-year-olds		
Mean (SD)	41.3 (12.4)	35.7 (8.9)
Range	24 – 67	20 – 51
5-year-olds		
Mean (SD)	44.0 (10.4)	40.1 (10.2)
Range	28 – 65	26 – 62
6-year-olds		
Mean (SD)	53.0 (12.2)	46.5 (9.8)
Range	34 – 81	25 – 77
Swedish-German bilinguals (N=46)		
Mean (SD)	43.6 (10.6)	40.2 (11.9)
Range	21 – 67	20 – 75
4-year-olds		
Mean (SD)	40.2 (7.8)	34.0 (8.6)
Range	24 – 60	22 – 51
5-year-olds		
Mean (SD)	38.7 (9.9)	38.5 (11.7)
Range	21 – 58	20 – 58
6-year-olds		
Mean (SD)	51.0 (9.5)	47.4 (11.4)
Range	39 – 67	33 – 75
Swedish-Turkish bilinguals (N=48)		
Mean (SD)	38.5 (12.1)	41.8 (12.7)
Range	11 – 67	23 – 76
4-year-olds		
Mean (SD)	33.8 (11.9)	36.4 (8.9)
Range	11 – 56	23 – 58
5-year-olds		
Mean (SD)	42.3 (12.9)	43.2 (15.5)
Range	15 – 67	23 – 76
6-year-olds		
Mean (SD)	39.1 (10.5)	45.6 (11.6)
Range	21 – 62	30 – 70

Table A3.6. Number of Different Words (NDW), German MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), Swedish-German bilinguals (N=46), by age group.

	MAIN1	**MAIN2**
Mean (SD)	44.9 (17.1)	38.8 (13.4)
Range	17 – 86	20 – 79
4-year-olds		
Mean (SD)	42.9 (14.0)	36.6 (12.2)
Range	17 – 68	23 – 69
5-year-olds		
Mean (SD)	40.3 (18.7)	33.3 (9.4)
Range	18 – 80	20 – 54
6-year-olds		
Mean (SD)	51.4 (16.9)	46.2 (15.0)
Range	30 – 86	24 – 79

Table A3.7. Number of Different Words (NDW), Turkish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), Swedish-Turkish bilinguals (N=48), by age group.

	MAIN1	**MAIN2**
Mean (SD)	35.4 (10.3)	40.1 (16.3)
Range	9 – 58	10 – 89
4-year-olds		
Mean (SD)	33.4 (11.7)	37.2 (16.9)
Range	22 – 58	14 – 84
5-year-olds		
Mean (SD)	34.5 (10.5)	40.6 (16.5)
Range	9 – 48	10 – 67
6-year-olds		
Mean (SD)	38.1 (8.4)	43.4 (15.9)
Range	25 – 53	24 – 89

Appendix 4 Character Introduction

Table A4.1. Total number of Swedish referring expressions used for introducing story characters, by age group and language group, MAIN1 (Cat/Dog), ENNI (A2/B2) and total.

	MAIN1	**ENNI**	**Total**
Monolinguals (N=72)			
4-year-olds	70	62	132
5-year-olds	66	67	133
6-year-olds	72	66	138
Total	**208**	**195**	**403**
Swedish-German bilinguals (N=46)			
4-year-olds	40	33	73
5-year-olds	43	38	81
6-year-olds	48	47	95
Total	**131**	**118**	**249**
Swedish-Turkish bilinguals (N=48)			
4-year-olds	43	33	76
5-year-olds	45	37	82
6-year-olds	48	40	88
Total	**136**	**110**	**246**
Total all groups	**475**	**423**	**898**

Table A4.2. Mean number of Swedish referring expressions per child, by age group and language group, MAIN1 (Cat/Dog), and ENNI (A2/B2).

	MAIN1	**ENNI**
Monolinguals (N=72)		
4-year-olds	2.9	2.6
5-year-olds	2.8	2.8
6-year-olds	3.0	2.8
Total	**2.9**	**2.7**
Swedish-German bilinguals (N=46)		
4-year-olds	2.9	2.5
5-year-olds	2.9	2.4
6-year-olds	3.0	2.9
Total	**2.9**	**2.6**
Swedish-Turkish bilinguals (N=48)		
4-year-olds	2.7	2.5
5-year-olds	2.8	2.5
6-year-olds	3.0	2.7
Total	**2.8**	**2.6**
Total all groups	**2.9**	**2.6**

Table A4.3. Proportions (%) of different types of referring expressions used for character introduction, Swedish MAIN1 (Cat/Dog), by age group, for the three language groups separated. (Raw figures in parentheses).

	Monolinguals (N=72)			**Swedish-German bilinguals (N=46)**			**Swedish-Turkish bilinguals (N=48)**		
	4yrs	**5yrs**	**6yrs**	**4yrs**	**5yrs**	**6yrs**	**4yrs**	**5yrs**	**6yrs**
Pronouns	11.4	10.6	1.4	10.0	2.3	4.2	16.3	11.1	6.3
	(8)	(7)	(1)	(4)	(1)	(2)	(7)	(5)	(3)
Bare Nouns	8.6	4.5	1.4	7.5	4.7	2.1	7.0	6.7	4.2
	(6)	(3)	(1)	(3)	(2)	(1)	(3)	(3)	(2)
Definite NPs	38.6	19.7	6.9	12.5	27.9	8.3	46.5	28.9	25.0
	(27)	(13)	(5)	(5)	(12)	(4)	(20)	(13)	(12)
Indefinite NPs	40.0	65.2	90.3	70.0	65.1	85.4	30.2	53.3	64.6
	(28)	(43)	(65)	(28)	(28)	(41)	(13)	(24)	(31)
Possessive NPs	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(1)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	(70)	**(66)**	**(72)**	**(40)**	**(43)**	**(48)**	**(43)**	**(45)**	**(48)**

Table A4.4. Proportions (%) of different types of referring expressions used for character introduction, Swedish ENNI (A2/B2), by age group, for the three language groups separated. (Raw figures in parentheses).

	Monolinguals (N=72)			**Swedish-German bilinguals (N=46)**			**Swedish-Turkish bilinguals (N=48)**		
	4yrs	**5yrs**	**6yrs**	**4yrs**	**5yrs**	**6yrs**	**4yrs**	**5yrs**	**6yrs**
Pronouns	35.5	19.4	10.6	15.2	21.1	8.5	36.4	32.4	25.0
	(22)	(13)	(7)	(5)	(8)	(4)	(12)	(12)	(10)
Bare Nouns	8.1	6.0	0.0	9.1	2.6	2.1	15.2	16.2	5.0
	(5)	(4)	(0)	(3)	(1)	(1)	(5)	(6)	(2)
Definite NPs	30.6	29.9	18.2	51.5	39.5	31.9	24.2	29.7	32.5
	(19)	(20)	(12)	(17)	(15)	(15)	(8)	(11)	(13)
Indefinite NPs	17.7	35.8	62.1	15.2	34.2	53.2	15.2	16.2	32.5
	(11)	(24)	(41)	(5)	(13)	(25)	(5)	(6)	(13)
Possessive NPs	4.8	6.0	6.1	6.1	2.6	2.1	9.1	5.4	2.5
	(3)	(4)	(4)	(2)	(1)	(1)	(3)	(2)	(1)
Proper Nouns	3.2	3.0	3.0	3.0	0.0	2.1	0.0	0.0	2.5
	(2)	(2)	(2)	(1)	(0)	(1)	(0)	(0)	(1)
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	(62)	**(67)**	**(66)**	**(33)**	**(38)**	**(47)**	**(33)**	**(37)**	**(40)**

Table A4.5. Summary of logistic regression model: Pronouns versus lexical NPs, Swedish MAIN1 (Cat/Dog) and ENNI (A2/B2), alternative coding of the variable language group.

Predictor	***β***	**SE**	**z (Wald)**	***p* value**
Intercept/Constant	-1.312	.127	107.482	< .001***
Narrative task: MAIN1 vs ENNI	-1.249	.210	35.188	< .001***
Age (1): 6 vs 4 & 5	-.914	.234	15.211	< .001***
Age (2): 5 vs 4	-.400	.226	3.134	.077
Language group (1): Swedish-Turkish vs Swedish-German & Swedish monolinguals	.702	.214	10.788	.001**
Language group (2): Swedish-German vs Swedish monolinguals	-.423	.265	2.561	.110
Model evaluation				
R^2 (Nagelkerke)	.129			
-2 Log likelihood	678.082			

Note. ** = p < .01, *** = p < .001. The second value of each predictor is the reference level for that predictor.

Table A4.6. Summary of logistic regression model: Fully appropriate NPs versus all other NPs, Swedish MAIN1 (Cat/Dog) and ENNI (A2/B2), alternative coding of the variable language group.

Predictor	***β***	**SE**	**z (Wald)**	***p* value**
Intercept/Constant	-.514	.109	22.043	< .001***
Narrative task: MAIN1 vs ENNI	1.106	.150	54.177	< .001***
Age (1): 6 vs 4 & 5	1.222	.162	57.219	< .001***
Age (2): 5 vs 4	.552	.185	8.938	.003**
Language group (1): Swedish-Turkish vs Swedish-German & Swedish monolinguals	-.856	.166	26.588	< .001***
Language group (2): Swedish-German vs Swedish monolinguals	-.020	.178	0.013	.910
Age (1) x Language group (1)	-.534	.343	2.426	.119
Age (1) x Language group (2)	-.790	.387	4.167	.041*
Age (2) x Language group (1)	.024	.412	0.003	.954
Age (2) x Language group (2)	-.830	.426	3.795	.051
Model evaluation				
R^2 (Nagelkerke)	.224			
-2 Log likelihood	1076.837			

Note. * = p < .05, ** = p < .01, *** = p < .001. The second value of each predictor is the reference level for that predictor.

Table A4.7. Number of referring expressions, by age group and narrative task, German MAIN1 (Cat/Dog), ENNI (A2/B2) and total, Swedish-German bilinguals (N=46).

	MAIN1	**ENNI**	**Total**
4-year-olds	39	33	72
5-year-olds	44	43	87
6-year-olds	48	41	89
Total	**131**	**117**	**248**

Table A4.8. Mean number of referring expressions per child by age group and narrative task, German MAIN1 (Cat/Dog), and ENNI (A2/B2), Swedish-German bilinguals (N=46).

	MAIN1	**ENNI**
4-year-olds	2.8	2.5
5-year-olds	2.8	2.7
6-year-olds	3.0	2.7
Total	**2.9**	**2.7**

Table A4.9. Proportions of different types of referring expressions used for character introduction, German MAIN1 (Cat/Dog) and ENNI (A2/B2), by age group, Swedish-German bilinguals (N=46). (Raw figures in parentheses).

	MAIN1			**ENNI**		
	4yrs	**5yrs**	**6yrs**	**4yrs**	**5yrs**	**6yrs**
Pronouns	15.4	4.5	2.1	15.2	11.6	7.3
	(6)	(2)	(1)	(5)	(5)	(3)
Bare Nouns	5.1	0.0	2.1	6.1	4.7	0.0
	(2)	(0)	(1)	(2)	(2)	(0)
Definite NPs	30.8	36.4	10.4	27.3	34.9	36.6
	(12)	(16)	(5)	(9)	(15)	(15)
Indefinite NPs	48.7	59.1	85.4	51.5	48.8	53.7
	(19)	(26)	(41)	(17)	(21)	(22)
Possessive NPs	0.0	0.0	0.0	0.0	0.0	2.4
	(0)	(0)	(0)	(0)	(0)	(1)
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	(39)	**(44)**	**(48)**	**(33)**	**(43)**	**(41)**

Appendix 5 Narrative macrostructure

Table A5.1. Swedish MAIN2 (Baby Birds/Baby Goats) total comprehension scores, pairwise comparisons between the language groups.

	MAIN2	**BG only**
Monolinguals vs Swedish-German	.867	.850
Monolinguals vs Swedish-Turkish	.001**	.006**
Swedish-German vs Swedish-Turkish	< .001***	.001**

Note. BG = Baby Goats, ** = p < .01, *** = p < .001.

Table A5.2. Swedish MAIN1 (Cat/Dog), Swedish, German and Turkish MAIN2 (Baby Birds/Baby Goats) total comprehension scores, pairwise comparisons between the age groups.

	SweMAIN1	**SweMAIN2**	**GerMAIN2**	**TurMAIN2**
4yrs vs 5yrs	.830	.031*	1.00	.198
4yrs vs 6yrs	< .001***	< .001***	.067	.017*
5yrs vs 6yrs	< .001***	< .001***	.095	.914

Note. * = p < .05, *** = p < .001.

Table A5.3. Swedish MAIN1 comprehension (Cat/Dog), percentage (%) correct answers on the different questions, by language group.

	Monolinguals (N = 72)	**Swedish-German bilinguals (N = 45)**	**Swedish-Turkish bilinguals (N = 46)**
D1. Episode 1 Goal	85	84	87
D2. Episode 1 IST	81	82	74
D3. Episode 1 IST rationale	78	73	63
D4. Episode 2 Goal	86	84	87
D5. Episode 2 IST	92	98	89
D6. Episode 2 IST rationale	82	98	80
D7. Episode 3 Goal	93	98	93
D8. Theory of Mind IST	82	91	74
D9. Theory of Mind IST rationale	78	84	70
D10. Overall plotline	29	42	29

Table A5.4. Swedish MAIN2 comprehension (Baby Birds/Baby Goats), percentage (%) correct on the different questions, by language group.

	Monolinguals (N = 72)	**Swedish-German bilinguals (N = 45)**	**Swedish-Turkish bilinguals (N = 46)**
D1. Episode 1 Goal	75	87	48
D2. Episode 1 IST	74	72	67
D3. Episode 1 IST rationale	63	57	35
D4. Episode 2 Goal	89	93	80
D5. Episode 2 IST	94	93	78
D6. Episode 2 IST rationale	86	85	52
D7. Episode 3 Goal	64	74	58
D8. Theory of Mind IST	42	50	49
D9. Theory of Mind IST rationale	26	43	29
D10. Overall plotline	75	72	49

Table A2.5. Swedish MAIN2 (Baby Birds/Baby Goats) goal comprehension, pair-wise comparisons between the language groups.

	p-value
Monolinguals vs Swedish-German	.146
Monolinguals vs Swedish-Turkish	.06
Swedish-German vs Swedish-Turkish	< .001***

Note. *** = p < .001.

Table A5.6. Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) total production scores, pairwise comparisons between the language groups.

	MAIN1	**MAIN2**
Monolinguals vs Swedish-German	.958	.848
Monolinguals vs Swedish-Turkish	< .001***	< .001***
Swedish-German vs Swedish-Turkish	< .001***	< .001***

Note. *** = p < .001.

Table A5.7. Swedish MAIN1 (Cat/Dog), Swedish, German and Turkish MAIN2 (Baby Birds/Baby Goats) total production scores, pairwise comparisons between the age groups.

	SweMAIN1	**SweMAIN2**	**GerMAIN2**	**TurMAIN2**
4yrs vs 5yrs	< .001***	.033*	1.00	.100
4yrs vs 6yrs	< .001***	< .001***	.008**	.054
5yrs vs 6yrs	< .001***	.002**	.018*	1.00

Note. * = p < .05, ** = p < .01, *** = p < .001.

Table A5.8. Swedish MAIN1 production (Cat/Dog), percentage (%) of the different components produced, by language and age group.

	Monolinguals (N = 72)				**Swedish-German bilinguals (N = 45)**				**Swedish-Turkish bilinguals (N = 48)**			
	4yrs	5yrs	6yrs	**Total**	4yrs	5yrs	6yrs	**Total**	4yrs	5yrs	6yrs	**Total**
A1. Setting: Time	4	38	33	**25**	0	13	38	**18**	13	0	0	**4**
A1. Setting: Place	0	17	33	**17**	0	20	38	**20**	0	0	6	**2**
A2. IST as IE (Ep1)	8	17	29	**18**	14	7	50	**24**	6	19	44	**23**
A3. Goal (Ep1)	13	25	29	**22**	50	40	50	**47**	31	38	38	**35**
A4. Attempt (Ep1)	54	79	83	**72**	50	67	75	**64**	13	56	63	**44**
A5. Outcome (Ep1)	75	75	96	**82**	71	80	94	**82**	31	56	63	**50**
A6. IST as R (Ep1)	4	8	8	**7**	7	20	0	**9**	0	6	0	**2**
A7. IST as IE (Ep2)	13	8	13	**11**	7	7	44	**20**	0	6	6	**4**
A8. Goal (Ep2)	4	13	8	**8**	14	7	25	**16**	0	31	6	**13**
A9. Attempt (Ep2)	58	79	100	**79**	71	53	75	**67**	13	19	63	**31**
A10. Outcome (Ep2)	75	75	96	**82**	71	67	94	**78**	75	75	94	**81**
A11. IST as R (Ep2)	4	4	4	**4**	7	0	6	**4**	13	13	13	**13**
A12. IST as IE (Ep3)	13	21	38	**24**	7	40	31	**27**	13	31	19	**21**
A13. Goal (Ep3)	8	21	8	**13**	14	7	19	**13**	6	31	19	**19**
A14. Attempt (Ep3)	38	58	63	**53**	36	53	63	**51**	6	19	44	**23**
A15. Outcome (Ep3)	75	88	92	**85**	86	93	94	**91**	50	69	94	**71**
A16. IST as R (Ep3)	4	4	0	**3**	7	7	6	**7**	0	0	0	**0**

Note. N = number of children, Ep1 = Episode 1, Ep2 = Episode 2, Ep3 = Episode 3, IST = internal state term, IE = initiating event, R = reaction.

Table A5.9. Swedish MAIN2 production (BB/BG), percentage (%) of the different components produced, by language and age group.

	Monolinguals (N = 72)				**Swedish-German bilinguals (N = 46)**				**Swedish-Turkish bilinguals (N = 48)**			
	4yrs	5yrs	6yrs	**Total**	4yrs	5yrs	6yrs	**Total**	4yrs	5yrs	6yrs	**Total**
A1. Setting: Time	8	17	25	**17**	0	19	31	**17**	13	0	13	**8**
A1. Setting: Place	13	17	42	**24**	29	31	31	**30**	31	13	19	**21**
A2. IST as IE (Ep1)	13	38	46	**32**	29	31	56	**39**	13	19	13	**15**
A3. Goal (Ep1)	4	4	29	**13**	14	13	0	**9**	13	19	25	**19**
A4. Attempt (Ep1)	42	63	79	**61**	29	56	94	**61**	31	38	56	**42**
A5. Outcome (Ep1)	38	54	67	**53**	36	56	88	**61**	19	38	50	**35**
A6. IST as R (Ep1)	4	4	4	**4**	0	0	0	**0**	13	0	0	**4**
A7. IST as IE (Ep2)	25	25	21	**24**	21	38	56	**39**	6	19	31	**19**
A8. Goal (Ep2)	25	42	21	**29**	57	25	56	**46**	31	44	38	**38**
A9. Attempt (Ep2)	63	83	83	**76**	50	69	81	**67**	44	56	56	**52**
A10. Outcome (Ep2)	54	54	71	**60**	36	56	69	**54**	31	38	88	**52**
A11. IST as R (Ep2)	17	13	13	**14**	7	19	19	**15**	13	6	0	**6**
A12. IST as IE (Ep3)	0	25	8	**11**	14	13	38	**22**	6	13	6	**8**
A13. Goal (Ep3)	8	8	0	**6**	0	6	13	**7**	0	6	0	**2**
A14. Attempt (Ep3)	92	92	96	**93**	93	94	100	**96**	44	50	88	**60**
A15. Outcome (Ep3)	50	71	96	**72**	71	69	100	**80**	31	44	38	**38**
A16. IST as R (Ep3)	42	25	25	**31**	21	19	19	**20**	19	19	31	**23**

Note. N = number of children, BB = Baby Birds, BG = Baby Goats, Ep1 = Episode 1, Ep2 = Episode 2, Ep3 = Episode 3, IST = internal state term, IE = initiating event, R = reaction.

Table A5.10. German/Turkish MAIN1 production (Cat/Dog), percentage (%) of the different components produced, by language and age group.

	Swedish-German bilinguals (N = 46)				**Swedish-Turkish bilinguals (N = 48)**			
	4yrs	5yrs	6yrs	**Total**	4yrs	5yrs	6yrs	**Total**
A1. Setting: Time	0	6	31	**13**	25	0	0	**8**
A1. Setting: Place	0	0	13	**4**	0	0	6	**2**
A2. IST as IE (Ep1)	7	19	6	**11**	6	31	19	**19**
A3. Goal (Ep1)	43	31	31	**35**	25	44	38	**35**
A4. Attempt (Ep1)	64	44	50	**52**	44	31	38	**38**
A5. Outcome (Ep1)	64	44	75	**61**	44	50	69	**54**
A6. IST as R (Ep1)	14	13	19	**15**	0	13	0	**4**
A7. IST as IE (Ep2)	0	19	25	**15**	13	13	6	**10**
A8. Goal (Ep2)	14	6	13	**11**	0	6	13	**6**
A9. Attempt (Ep2)	36	50	69	**52**	13	25	19	**19**
A10. Outcome (Ep2)	64	69	69	**67**	56	75	81	**71**
A11. IST as R (Ep2)	14	13	6	**11**	0	6	13	**6**
A12. IST as IE (Ep3)	7	25	50	**28**	6	31	44	**27**
A13. Goal (Ep3)	14	6	13	**11**	6	19	13	**13**
A14. Attempt (Ep3)	50	56	56	**54**	6	19	31	**19**
A15. Outcome (Ep3)	79	81	100	**87**	50	75	75	**67**
A16. IST as R (Ep3)	0	0	6	**2**	0	0	0	**0**

Note. N = number of children, Ep1 = Episode 1, Ep2 = Episode 2, Ep3 = Episode 3, IST = internal state term, IE = initiating event, R = reaction.

Table A5.11. German/Turkish MAIN2 production (Baby Birds/Baby Goats), percentage (%) of the different components produced, by language and age group.

	Swedish-German bilinguals (N = 46)				Swedish-Turkish bilinguals (N = 48)			
	4yrs	5yrs	6yrs	**Total**	4yrs	5yrs	6yrs	**Total**
A1. Setting: Time	0	6	25	**11**	19	0	6	**8**
A1. Setting: Place	21	38	44	**35**	13	25	25	**21**
A2. IST as IE (Ep1)	36	19	38	**30**	13	25	38	**25**
A3. Goal (Ep1)	0	6	31	**13**	6	6	31	**15**
A4. Attempt (Ep1)	71	63	63	**65**	19	19	50	**29**
A5. Outcome (Ep1)	14	38	56	**37**	13	63	63	**46**
A6. IST as R (Ep1)	0	0	0	**0**	0	0	6	**2**
A7. IST as IE (Ep2)	0	19	25	**15**	25	38	13	**25**
A8. Goal (Ep2)	43	31	38	**37**	38	63	50	**50**
A9. Attempt (Ep2)	64	44	69	**59**	25	31	44	**33**
A10. Outcome (Ep2)	36	50	44	**43**	25	50	56	**44**
A11. IST as R (Ep2)	0	0	0	**0**	19	6	6	**10**
A12. IST as IE (Ep3)	7	19	56	**28**	13	31	13	**19**
A13. Goal (Ep3)	0	6	6	**4**	0	0	0	**0**
A14. Attempt (Ep3)	79	75	88	**80**	44	81	69	**65**
A15. Outcome (Ep3)	50	50	88	**63**	31	63	56	**50**
A16. IST as R (Ep3)	7	0	38	**15**	13	25	25	**21**

Note. N = number of children, Ep1 = Episode 1, Ep2 = Episode 2, Ep3 = Episode 3, IST = internal state term, IE = initiating event, R = reaction.

Table A5.12. Swedish MAIN1 (Cat/Dog), types of components, pairwise comparisons between the language groups.

	Settings	**Goals**	**Attempts**	**Outcomes**
Monolinguals vs Swedish-German	1.00	.022	.540	1.00
Monolinguals vs Swedish-Turkish	.003**	.239	<.001***	.005**
Swedish-German vs Swedish-Turkish	.016*	1.00	<.001***	.006**

Note. * = p < .05, ** = p < .01, *** = p < .001.

Table A5.13. Swedish MAIN1 (Cat/Dog), types of components, pairwise comparisons between the age groups.

	Settings	**ISTs as IE**	**Attempts**	**Outcomes**
4yrs vs 5yrs	.062	.247	.010*	1.00
4yrs vs 6yrs	<.001***	<.001***	<.001***	<.001***
5yrs vs 6yrs	.085	.042*	.028*	.004**

Note. * = p < .05, ** = p < .01, *** = p < .001.

Table A5.14. Swedish MAIN2 (Baby Birds/Baby Goats), types of components, pairwise comparisons between the language groups.

	ISTs as IE	Attempts	Outcomes
Monolinguals vs Swedish-German	.305	1.00	1.00
Monolinguals vs Swedish-Turkish	.313	.001**	.007**
Swedish-German vs Swedish-Turkish	.009**	.011*	.003**

Note. * = p < .05, ** = p < .01.

Table A5.15. Swedish MAIN2 (Baby Birds/Baby Goats), types of components, pairwise comparisons between the age groups.

	Settings	IST as IE	Attempts	Outcomes
4yrs vs 5yrs	1.00	.056	.05	.200
4yrs vs 6yrs	.064	.006**	<.001***	<.001***
5yrs vs 6yrs	.089	1.00	.169	.002**

Note. ** = p < .01, *** = p < .001.

Table A5.16. Summary of logistic regression model: Macrostructural complexity, sequence versus no sequence, Swedish MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats), alternative coding of the variable language group.

Predictor	***β***	***SE***	***z* (Wald)**	***p* value**
Intercept/Constant	.192	.071	7.335	.007**
Age (1): 6 vs 4 & 5	1.292	.154	70.430	< .001***
Age (2): 5 vs 4	.687	.164	17.493	< .001***
Language group (1): Swedish-Turkish vs Swedish-German & Swedish mono-linguals	-1.207	.156	60.040	< .001***
Language group (2): Swedish-German vs Swedish monolinguals	.135	.169	.642	.423
Model evaluation				
R^2 (Nagelkerke)	.192			
-2 Log likelihood	1211.871			

Note. ** = p < .01, *** = p < .001. The second value of each predictor is the reference level for that predictor.

Appendix 6 Linking vocabulary and character introduction to macrostructure

Table A6.1. Swedish and German MAIN1 (Cat/Dog) macrostructure score and the number of story characters introduced with fully appropriate referring expressions (FAE), pairwise comparisons.

	Swedish	**German**
0 FAE vs 1 FAE	.006**	1.00
0 FAE vs 2 FAE	< .001***	.251
0 FAE vs 3 FAE	< .001***	.047*
1 FAE vs 2 FAE	.966	.606
1 FAE vs 3 FAE	.003**	.100
2 FAE vs 3 FAE	.215	1.00

Note. * = p < .05, ** = p < .01, *** = p < .001.

Table A6.2. Swedish MAIN1 (Cat/Dog) macrostructure score and the number of story characters introduced with lexical NPs (lexNP), pairwise comparisons.

	Swedish
1 lexNP vs 2 lexNP	.552
1 lexNP vs 3 lexNP	< .001***
2 lexNP vs 3 lexNP	< .001***

Note. *** = p < .001.

References

Aarssen, J. (1996). *Relating events in two languages: Acquisition of cohesive devices by Turkish-Dutch bilingual children at school age*. Tilburg: Tilburg University Press.

Akinci, M.-A., Jisa, H., & Kern, S. (2001). Influence of L1 Turkish on L2 French narratives. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (pp. 189–208). Amsterdam: John Benjamins.

Akoğlu, G., & Yağmur, K. (2016). First-language skills of bilingual Turkish immigrant children growing up in a Dutch submersion context. *International Journal of Bilingual Education and Bilingualism*, *19*(6), 706–721.

Aksu-Koç, A., & Nicolopoulou, A. (2015). Character reference in young children's narratives: A crosslinguistic comparison of English, Greek, and Turkish. *Lingua*, *155*, 62–84.

Allen, S., Skarabela, B., & Hughes, M. (2008). Using corpora to examine discourse effects in syntax. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 99–137). Amsterdam: John Benjamins.

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, *37*(1), 165–193.

Altman, C., Goldstein, T., & Armon-Lotem, S. (2017). Quantitative and qualitative differences in the lexical knowledge of monolingual and bilingual children on the LITMUS-CLT task. *Clinical Linguistics & Phonetics*, *31*(11–12), 931–954.

Álvarez, E. (2003). Character introduction in two languages: Its development in the stories of a Spanish-English bilingual child age 6;11-10;11. *Bilingualism: Language and Cognition*, *6*(3), 227–243.

Andreou, M., Knopp, E., Bongartz, C., & Tsimpli, I. M. (2015). Character reference in Greek-German bilingual children's narratives. *EUROSLA Yearbook*, *15*, 1–40.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527.

Arnoldsson, S., & Aronsson, B. (2013). *Femåringar berättar på svenska och engelska: Referensdata och jämförelse med 6- och 7-åringar* [Five-year-olds tell stories in Swedish and English: Reference data and comparisons with 6- and 7-year-olds] (Unpublished MA thesis). Uppsala University, Uppsala.

Bamberg, M., & Damrad-Frye, R. (1991). On the ability to provide evaluative comments: further explorations of children's narrative competencies. *Journal of Child Language*, *18*(3), 689–710.

Barron-Hauwaert, S. (2004). *Language strategies for bilingual families: the one-parent-one-language approach*. Clevedon: Multilingual Matters.

Berglund, E., & Eriksson, M. (2000). Communicative development in Swedish children 16-28 months old: The Swedish Early Communicative Development Inventory-Words and Sentences. *Scandinavian Journal of Psychology*, *41*(2), 133–144.

Berman, R. A. (1988). On the ability to relate events in narrative. *Discourse Processes*, *11*(4), 469–497.

Berman, R. A., & Slobin, D. I. (1994a). Different ways of relating events: introduction to the study. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: a crosslinguistic developmental study* (pp. 1–16). Hillsdale, NJ: Lawrence Erlbaum.

Berman, R. A., & Slobin, D. I. (1994b). Narrative structure. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: a crosslinguistic developmental study* (pp. 39–84). Hillsdale, NJ: Lawrence Erlbaum.

Berman, R. A., & Slobin, D. I. (Eds.). (1994c). *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.

Bernardini, P., & Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: The Ivy hypothesis. *Bilingualism: Language and Cognition*, *7*(1), 49–69.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531.

Bohnacker, U. (2003). Nominal phrases. In G. Josefsson, C. Platzack, & G. Håkansson (Eds.), *The acquisition of Swedish grammar* (pp. 195–260). Amsterdam: John Benjamins.

Bohnacker, U. (2007). The role of input frequency in article acquisition in early child Swedish. In I. Gülzow & N. Gagarina (Eds.), *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept* (pp. 51–82). Berlin/New York: Mouton de Gruyter.

Bohnacker, U. (2013). *Language impairment or typical language development? Developing methods for linguistic assessment of bilingual children in Sweden* Grant proposal to the Swedish Research Council, March 2013 [unpublished research programme].

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, *37*(1), 19–48.

Bohnacker, U., & Lindgren, J. (submitted). MAIN story comprehension: What can we expect of a typically developing child?

Bohnacker, U., Lindgren, J., & Öztekin, B. (2016). Turkish- and German-speaking bilingual 4-to-6-year-olds living in Sweden: Effects of age, SES and home language input on vocabulary production. *Journal of Home Language Research*, *1*, 17–40.

Bos, P. (1997). *Development of bilingualism: A study of school-age Moroccan children in the Netherlands*. Tilburg: Tilburg University Press.

Boudreau, D. (2007). Narrative abilities in children with language impairments. In R. Paul (Ed.), *Language disorders from a developmental perspective: essays in honor of Robin S. Chapman* (pp. 331–356). Mahwah, NJ: Lawrence Erlbaum Associates.

Brookshire, R. H., & McNeil, M. R. (2015). *Introduction to neurogenic communication disorders* (Eighth edition). St. Louis, MO: Elsevier.

Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Brusewitz, K., & Gómez-Ortega, M. (2005). *Boston Naming Test på svenska: normativa data för barn och ungdomar* [Boston Naming Test in Swedish: normative data for children and adolescents] (Unpublished MA-thesis). Karolinska Institutet, Stockholm.

Brusewitz, K., & Tallberg, I.-M. (2010). The Boston Naming Test and Swedish children: Normative data and response analysis. *European Journal of Developmental Psychology*, *7*(2), 265–280.

Buac, M., Gross, M., & Kaushanskaya, M. (2014). The role of primary caregiver vocabulary knowledge in the development of bilingual children's vocabulary skills. *Journal of Speech Language and Hearing Research*, *57*(5), 1804–1816.

Burris, S. E., & Brown, D. D. (2014). When all children comprehend: increasing the external validity of narrative comprehension development research. *Frontiers in Psychology*, *5*(Article 168).

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 25–55). New York, NY: Academic Press.

Chen, L., & Lei, J. (2013). The production of referring expressions in oral narratives of Chinese–English bilingual speakers and monolingual peers. *Child Language Teaching and Therapy*, *29*(1), 41–55.

Chen, L., & Pan, N. (2009). Development of English referring expressions in the narratives of Chinese–English bilinguals. *Bilingualism: Language and Cognition*, *12*(4), 429–445.

Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 123–148). Bristol: Multilingual Matters.

Cobo-Lewis, A. B., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002a). Effects of bilingualism and bilingual education on oral and written English skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 64–97). Clevedon: Multilingual Matters.

Cobo-Lewis, A. B., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002b). Effects of bilingualism and bilingual education on oral and written Spanish skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 98–117). Clevedon: Multilingual Matters.

Colozzo, P., & Whitely, C. (2014). Keeping track of characters: Factors affecting referential adequacy in children's narratives. *First Language*, *34*(2), 155–177.

Cox Eriksson, C. (2014). *Children's vocabulary development: the role of parental input, vocabulary composition and early communicative skills* (PhD Thesis). Stockholm University, Stockholm.

Dale, P. S. (2007). *MacArthur-Bates Communicative Development Inventory - III*. Baltimore, MD: Brookes.

De Cat, C. (2013). Egocentric definiteness errors and perspective evaluation in preschool children. *Journal of Pragmatics*, *56*, 58–69.

De Cat, C. (2015). The cognitive underpinnings of referential abilities. In L. Serratrice & S. E. M. Allen (Eds.), *The acquisition of reference* (pp. 263–283). Amsterdam: John Benjamins.

De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol: Multilingual Matters.

Dickinson, D., & Tabors, P. (2001). *Beginning literacy with language*. Baltimore, MD: Brookes.

Dijkstra, J., Kuiken, F., Jorna, R. J., & Klinkenberg, E. L. (2016). The role of majority and minority language input in the early development of a bilingual vocabulary. *Bilingualism: Language and Cognition*, *19*(1), 191–205.

Dillström, S., & Kesti, M. (2009). *Kan dold språkförståelse på andraspråket bli synlig vid återberättande på förstaspråket? Språkförståelse och språkproduktion i förskolebarns berättelser av Bussagan (BST) på svenska och arabiska* [Can hidden language comprehension in the second language become visible when retelling in the first language? Language comprehension and production in preschool children's Bus Story narratives in Swedish and Arabic] (Unpublished MA thesis). Uppsala University, Uppsala.

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *International Journal of Language & Communication Disorders*, *43*(2), 201–218.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT III. Peabody Picture Vocabulary Test*. Third Edition. Circle Pines, MN: American Guidance Service.

Emslie, H. C., & Stevenson, R. J. (1981). Pre-school children's use of the articles in definite and indefinite referring expressions. *Journal of Child Language*, *8*(2), 313–328.

Eriksson, M. (2017). The Swedish Communicative Development Inventory III: Parent reports on language in preschool children. *International Journal of Behavioral Development*, *41*(5), 647–654.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., Pethick, S., & Reilly, J.S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics* (4th edition). Los Angeles, CA: Sage.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: language and task effects. *Language, Speech, and Hearing Services in Schools*, *35*(2), 155–168.

Finnstedt, L. (2013). *Referentiell kohesion och temporal förankring i barns narrativer: Jämförelse mellan enspråkiga svensktalande och flerspråkiga engelsksvensktalande barn* [Referential cohesion and temporal anchoring in children's narratives: A comparison of monolingual Swedish-speaking and multilingual English-Swedish-speaking children] (Unpublished MA thesis). Uppsala University, Uppsala.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Fraurud, K. (1996). Cognitive ontology and NP form. In T. Fretheim & J. K. Gundel (Eds.), *Reference and Referent Accessibility* (pp. 65–88). Amsterdam: John Benjamins.

Frithiofsson, K., & Öberg, L. (2012). *Skillnader mellan enspråkiga och flerspråkiga barns berättande: Narrativer och lexikon på svenska och engelska* [Differences between mono- and bilingual children's narrations: Narratives and lexicon in Swedish and English] (Unpublished MA-thesis). Uppsala University, Uppsala.

Frylmark, A. (2006). *Bedömning Av Språk B.A.S. 5½-7½ år*. Märsta: OrdAF.

Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: Rasskaz and Erzählung. *Applied Psycholinguistics*, *37*(1), 91–122.

Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., Golcher, F., & Walters, J. (2014). Age, input quantity and their effect on linguistic performance in the home and societal language among Russian-German and Russian-Hebrew preschool children. In R. K. Silbereisen, P. F. Titzmann, & Y. Shovit (Eds.), *The challenges of diaspora migration: Interdisciplinary perspectives on Israel and Germany* (pp. 63–82). Farnham: Ashgate.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Asssessment Instrument for Narratives. *ZAS Papers in Linguistics*, *56*, 1–156.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–269). Bristol: Multilingual Matters.

Gagarina, N., Klop, D., Tsimpli, I. M., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, *37*(1), 11–17.

Ganuza, N., & Hedman, C. (2017). The impact of mother tongue instruction on the development of biliteracy: Evidence from Somali–Swedish bilinguals. *Applied Linguistics*, First view, 1–25.

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, *12*(2), 213–237.

Gathercole, V. C. M., Thomas, E. M., Roberts, E. J., Hughes, C. O., & Hughes, E. K. (2013). Why assessment needs to take exposure into account: Vocabulary and grammatical abilities in bilingual children. In V. C. M. Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 20–55). Bristol: Multilingual Matters.

Gatt, D., Attard, D., Łuniewska, M., & Haman, E. (2017). The effects of bilingual status on lexical comprehension and production in Maltese five-year-old children: A LITMUS-CLT study. *Clinical Linguistics & Phonetics*, *31*(11–12), 844–873.

Givón, T. (1983). *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

Granfeldt, J. (2000). The acquisition of the determiner phrase in bilingual and second language French. *Bilingualism: Language and Cognition*, *3*(3), 263–280.

Greenhalgh, K. S., & Strong, C. J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools*, *32*(2), 114–125.

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language*, *24*(2), 123–147.

Guidelines for scoring CLT (version March 2018). Bohnacker U., in collaboration with Dillström, S., Lindgren, J. & Öztekin, B. Guidelines developed for the BiLI-TAS project, Department of Linguistics & Philology, Uppsala University.

Guidelines for scoring nacrostructure in MAIN (version March 2018). Bohnacker, U., in collaboration with Lindgren, J., Dillström, S., Koltay, K. & Öztekin, B. Guidelines developed for the BiLI-TAS project, Department of Linguistics & Philology, Uppsala University.

Guidelines for the transcription of MAIN (version March 2018). Bohnacker, U., in collaboration with Lindgren, J., Dillström, S. & Öztekin, B. Guidelines developed for the BiLI-TAS project. Department of Linguistics & Philology, Uppsala University.

Guiraud, P. (1959). *Problèmes et méthodes de la statistique linguistique*. [Problems and methods in statistics for linguistics]. Dordrecht: Reidel.

Gülzow, I. (in preparation). Establishing discourse referents: Indefinite noun phrases in German preschooler's narratives. Unpublished Manuscript.
Gülzow, I., & Gagarina, N. (2007). Noun phrases, pronouns and anaphoric reference in young children narratives. *ZAS Papers in Linguistics*, *48*, 203–223.
Gundel, J. K. (2010). Reference and accessibility from a Givenness Hierarchy perspective. *International Review of Pragmatics*, *2*(2), 148–168.
Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.
Gundel, J. K., & Johnson, K. (2013). Children's use of referring expressions in spontaneous discourse: Implications for theory of mind development. *Journal of Pragmatics*, *56*, 43–57.
Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, *13*(2), 175–197.
Haessig, A., & Tuvås, L. (2013). *Narrativ förmåga på franska och svenska: En komparativ studie om enspråkiga och simultant flerspråkiga barn* [Narrative ability in French and Swedish: A comparative study of monolingual and simultaneous bilingual children] (Unpublished MA-thesis). Uppsala University, Uppsala.
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
Haman, E., Łuniewska, M., Hansen, P., Simonsen, H.G., Chiat, S., Bjekić, J., Blažienė, A., Chyl, K., Dabašinskienė, I., Engel de Abreu, P., Gagarina, N., Gavarró, A., Håkansson, G., Harel, E., Holm, E., Kapalková, S., Kunnari, S., Levorato, C., Lindgren, J., Mieszkowska, K., Montes Salarich, L., Potgieter, A., Ribu, I., Ringblom, N., Rinker, T., Roch, M., Slančová, D., Southwood, F., Tedeschi, R., Tuncer, A.M., Ünal-Logacev, Ö., Vuksanović, J., & Armon-Lotem, S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from cross-linguistic lexical tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, *31*(11–12), 818–843.
Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–239). Bristol: Multilingual Matters.
Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Miękisz, A., Kacprzak, A., Banasik, N., & Foryś-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers in Psychology*, *8*(Article 1444).
Härdelin, S., & Naylor, I. (2012). *Barn berättar på engelska och svenska: En explorativ studie om narrativer hos flerspråkiga barn i tidig skolålder* [Children tell in English and Swedish: An explorative study of narratives in multilingual children in the early school age] (Unpublished MA thesis). Uppsala University, Uppsala.
Harris, P. (1996). Desires, beliefs, and language. In P. Carruthers & P. Smith (Eds.), *Theories of theories of mind* (pp. 200–220). Cambridge: Cambridge University Press.
Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul Brookes.
Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology*, *19*(2), 154–166.

Hickmann, M. (2002). *Children's discourse: Person, space and time across languages*. Cambridge: Cambridge University Press.
Hickmann, M. (2004). Coherence, cohesion, and context: Some comparative perspectives in narrative development. In S. Strömqvist & L. T. Verhoeven (Eds.), *Relating events in narrative. Vol. 2, Typological and contextual perspectives* (pp. 281–306). Hillsdale, NJ: Lawrence Erlbaum.
Hickmann, M., Hendriks, H., Roland, F., & Liang, J. (1996). The marking of new information in children's narratives: a comparison of English, French, German and Mandarin Chinese. *Journal of Child Language*, *23*(3), 591–619.
Hoff, E., Core, C., Place, S., Rumiche, R., SeñOr, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27.
Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, *29*(4), 433–444.
Holmström, K. (2015). *Lexikal organisation hos en- och flerspråkiga skolbarn med språkstörning* [Lexical organization in mono- and multilingual school children with language impairment] (Unpublished PhD thesis). Lund University, Lund.
Hudson, J. A., & Shapiro, L. R. (1991). From knowing to telling: The development of children's scripts, stories, and personal narratives. In A. McCabe & C. Peterson (Eds.), *Developing narrative structure* (pp. 89–136). Hillsdale, NJ: Lawrence Erlbaum.
Hughes, M. E., & Allen, S. E. M. (2013). The effect of individual discourse-pragmatic features on referential choice in child English. *Journal of Pragmatics*, *56*, 15–30.
Hunt, K. (1965). *Grammatical Structures written at three grade levels*. Champaign, IL: National Council of Teachers of English (NCTE).
Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*(1), 58–74.
Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition*, *18*(4), 737–752.
Josefsson, G., Platzack, C., & Håkansson, G. (Eds.). (2003). *The acquisition of Swedish grammar*. Amsterdam: John Benjamins.
Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The Index of Narrative Microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology*, *15*(2), 177–191.
Kail, M., & Hickmann, M. (1992). French children's ability to introduce referents in narratives as a function of mutual knowledge. *First Language*, *12*(34), 73–94.
Kail, M., & Sanchez y Lopez, I. (1997). Referent introductions in Spanish narratives as a function of contextual constraints: a crosslinguistic perspective. *First Language*, *17*(51), 103–130.
Kalnak, N., Peyrard-Janvid, M., Sahlén, B., & Forssberg, H. (2012). Family history interview of a broad phenotype in specific language impairment and matched controls. *Genes, Brain and Behavior*, *11*, 921–927.
Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, *37*(1), 145–164.

Kapalková, S., & Slančová, D. (2017). The vocabulary profile of Slovak children with primary language impairment compared to typically developing Slovak children measured by LITMUS-CLT. *Clinical Linguistics & Phonetics*, *31*(11–12), 893–909.

Khoury Aouad Saliby, C., dos Santos, C., Kouba Hreich, E., & Messarra, C. (2017). Assessing Lebanese bilingual children: The use of Cross-linguistic Lexical Tasks in Lebanese Arabic. *Clinical Linguistics & Phonetics*, *31*(11–12), 874–892.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, *43*(6), 456–473.

Koivistoinen, J. (2012). *Att berätta på ryska och på svenska: Narrativ förmåga som instrument för språklig bedömning av flerspråkiga barn* [To tell a story in Russian and Swedish: Narrative ability as an instrument for language assessment of multilingual children] (Unpublished MA thesis). Uppsala University, Uppsala.

Krüger Vahlquist, M. (2012). *Snabb benämningsförmåga enligt Ordracet: normer samt felsvarsanalys för 6-10-åringar* [Rapid naming using the test Ordracet: norms and error analysis for 6-10-year-olds] (Unpublished MA thesis). Uppsala University, Uppsala.

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, *37*(1), 123–144.

Küntay, A. C. (2002). Development of the Expression of indefiniteness: Presenting new referents in Turkish picture-series stories. *Discourse Processes*, *33*(1), 77–101.

Kupisch, T., Anderssen, M., Bohnacker, U., & Snape, N. (2009). Article acquisition in German, English, Norwegian and Swedish. In W. Leow, H. Campos, & D. Lardiere (Eds.), *Little words: Their history, phonology, syntax, semantics, pragmatics, and acquisition* (pp. 223–235). Washington, DC: Georgetown University Press.

Labov, W. (1972). *Language in the inner city: studies in the black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington.

Leback, A., & Nilsson, L. (2012). *Simultant flerspråkiga barns narrativa förmågor: Referensdata för svensk-engelska barn* [Simultaneous multilingual children's narrative abilities: Reference data for Swedish-English children] (Unpublished MA thesis). Uppsala University, Uppsala.

Leonard, L. B. (2000). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leseman, P. P. M. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Multilingual and Multicultural Development*, *21*(2), 93–112.

Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind.' *Psychological Review*, *94*(4), 412–426.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind.' *Trends in Cognitive Sciences*, *8*(12), 528–533.

Liles, B. Z. (1993). Narrative discourse in children with language disorder and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research*, *36*(5), 868–882.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, *38*(2), 415–425.

Lindberg, I., & Johansson Kokkinakis, S. (2007). *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år* [OrdiL – a corpus-based survey of the vocabulary in study material for the later years of the primary school] (Rapporter om svenska som andraspråk No. 8). Gothenburg: University of Gothenburg.

Lindgren, J. (2017). Ordförrådsutveckling hos förskolebarn med svenska som modersmål: En studie med Cross-linguistic Lexical Task (CLT) [Vocabulary development in preschool children with Swedish as mother tongue: A study with Cross-linguistic Lexical Task (CLT)]. In E. Sköldberg, M. Andréasson, H. Adamsson Eryd, F. Lindahl, J. Prentice, S. Lindström, & M. Sandberg (Eds.), *Svenskans beskrivning 35: förhandlingar vid trettiofemte sammankomsten Göteborg 11-13 maj 2016* (pp. 161–174). Gothenburg: University of Gothenburg.

Lindgren, J. (2018). Effects of age and stimulus material on character introductions of Swedish-speaking four- to six-year-olds. *Journal of Child Language*, *45*(2), 526–540.

Lindgren, J., & Bohnacker, U. (submitted). Vocabulary in closely related languages: Age and word type effects in bilingual Swedish-German preschool children.

Lindgren, J., & Vogels, J. (submitted). Referential cohesion in Swedish preschool children's narratives.

Loban, W. (1963). *The language of elementary school children*. Champaign, IL: National Council of Teachers of English (NCTE).

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (Third Edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mäkinen, L. (2014). *Narrative language in typically developing children, children with specific language impairment and children with autism spectrum disorder* (PhD Thesis). University of Oulu, Oulu.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*(3), 403–422.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, *37*(1), 69–89.

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press.

McCabe, A. (1996). Evaluating narrative discourse skills. In K. Cole, P. S. Dale, & D. Thal (Eds.), *Assessment of communication and language* (pp. 121–142). Baltimore, MD: Paul Brookes.

McCabe, A., & Rollins, P. (1994). Assessment of preschool narrative skills: Prerequisite for literacy. *American Journal of Speech-Language Pathology*, *3*, 45–56.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished PhD thesis). University of Memphis, Memphis.

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic Status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, *8*(Article 1442).

Merritt, D. D., & Liles, B. Z. (1989). Narrative analysis: clinical applications of story generation and story retelling. *The Journal of Speech and Hearing Disorders*, *54*(3), 438–447.

Mikoczy, T., & Nyman, A. (2008). *Lexikal förmåga hos flerspråkiga svensk-arabiska barn i skolår 4: En undersökning med PPVT-III och Kent-Rosanoffs associationstest* [Lexical ability in multilingual Swedish-Arabic children in Grade 4: A study with PPVT-III and Kent-Rosanoff's association test] (Unpublished MA thesis). Lund University, Lund.

Miniscalco, C., Hagberg, B., Kadesjö, B., Westerlund, M., & Gillberg, C. (2007). Narrative skills, cognitive profiles and neuropsychiatric disorders in 7–8-year-old children with late developing language. *International Journal of Language & Communication Disorders*, *42*(6), 665–681.

Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language Speech and Hearing Services in Schools*, *34*(4), 332–342.

Nakamura, K. (1993). Referential structure in Japanese children's narratives: The acquisition of wa and ga. *Japanese/Korean Linguistics*, *3*, 84–99.

Namei, S. (2002). *The bilingual lexicon from a developmental perspective: a word association study of Persian-Swedish bilinguals* (PhD Thesis). Stockholm University, Stockholm.

Newman, R. M., & McGregor, K. (2006). Teachers and laypersons discern quality differences between narratives produced by children with and without SLI. *Journal of Speech Language and Hearing Research*, *49*(5), 1022–1036.

Nilsson, A., & Vikström, L. (1997). *Normalspråkiga fyra- och femåringars förmåga att strukturera och förstå berättelser* [Typically-developing four- and five-year-olds ability to structure and understand narratives] (Unpublished MA thesis). Lund University, Lund.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313.

Nordqvist, Å. (2001). *Speech about speech: A developmental study on form and function of direct and indirect speech* (PhD Thesis). University of Gothenburg, Gothenburg.

Olsson, A. (2013). *Tvåspråkiga barn berättar på ryska och svenska: En explorativ studie om narrativ förmåga hos 6–7-åringar* [Bilingual children tell in Russian and Swedish: An explorative study about narrative ability in 6–7-year-olds] (Unpublished MA thesis). Uppsala University, Uppsala.

Öztekin, B. (in preparation). *Typical and atypical language development in bilingual Turkish-Swedish preschool children* (PhD Thesis in preparation). Uppsala University, Uppsala.

Parkvall, M. (2015). *Sveriges språk i siffror: vilka språk talas och av hur många?* [Sweden's languages in numbers: which languages are spoken and by how many speakers?]. Stockholm: Språkrådet & Morfem.

Pavlenko, A. (2006). Narrative competence in a second language. In H. D. Weger-Guntharp, H. Byrnes, & K. A. Sprang (Eds.), *Educating for advanced foreign language capacities: Constructs, curriculum, instruction, assessment* (pp. 105–117). Washington, DC: Georgetown University Press.

Pavlenko, A. (2008). Narrative analysis. In L. Wei & M. G. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 311–325). Malden, MA: Blackwell.

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon: Multilingual Matters.

Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*(3), 399–410.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum Press.

Potgieter, A. P., & Southwood, F. (2016). A comparison of proficiency levels in 4-year-old monolingual and trilingual speakers of Afrikaans, isiXhosa and South African English across SES boundaries, using LITMUS-CLT. *Clinical Linguistics & Phonetics*, *30*(2), 87–100.

Prevoo, M. J. L., Malda, M., Mesman, J., Emmen, R. A. G., Yeniad, N., Van Ijzendoorn, M. H., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: different pathways for host and ethnic language. *Journal of Child Language*, *41*(5), 963–984.

Reichardt, V. (2014). *Wie führen Kinder Referenten ein? Eine Untersuchung der Verwendung deutscher Indefinita durch russisch-deutsch bilinguale Kinder verschiedener Altersgruppen* [How do children introduce referents? A study of the use of German indefinites by Russian-German bilingual children in different age groups] (Unpublished BA thesis). Humboldt-Universität zu Berlin, Berlin.

Reichardt, V. (2016). *Det var en gång en hund. Eine Untersuchung der Verwendung angemessener referentieller Ausdrücke zur Einführung von Referenten in Narrativen durch deutsch-schwedisch bilinguale Kinder im Alter von 4 und 6 Jahren* [Once upon a time, there was a dog. A study of the use of appropriate referring expressions to introduce referents in German-Swedish bilingual children aged 4 and 6] (Unpublished MA thesis). Humboldt-Universität zu Berlin, Berlin.

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language*, *88*(2), 229–247.

Renfrew, C. E. (1969). *Bus Story Test: A Test of Narrative Speech*. Bicester: Winslow Press.

Reuterskiöld, C., Hansson, K., & Sahlén, B. (2011). Narrative skills in Swedish children with language impairment. *Journal of Communication Disorders*, *44*(6), 733–744.

Reuterskiöld Wagner, C., Sahlén, B., & Nettelbladt, U. (1999). What's the story? Narration and comprehension in Swedish preschool children with language impairment. *Child Language Teaching and Therapy*, *15*(2), 83–93.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, *37*(1), 49–67.

Rosén, C. (2008). *„Warum klingt das nicht deutsch?": Probleme der Informationsstrukturierung in deutschen Texten schwedischer Schüler und Studenten* ["Why doesn't it sound German?": Problems with information structure in the German texts of Swedish pupils and students]. Saarbrücken: VDM-Verlag Müller.

Salameh, E.-K. (2011). Lexikal utveckling på svenska och arabiska vid tvåspråkig undervisning [Lexical development in Swedish and Arabic in bilingual education]. In L. Bergman, I. Ericsson, N. Hartsmar, L. Lang, B. Liljefors Persson, & C. Ljungberg (Eds.), *EDUCARE 3. Tema: Tvåspråkig undervisning på svenska och arabiska i mångkulturella storstadsskolor* (pp. 205–225). Malmö: Malmö Högskola.

Salameh, E.-K., Nettelbladt, U., & Gullberg, B. (2002). Risk factors for language impairment in Swedish bilingual and monolingual children relative to severity. *Acta Paediatrica*, *91*(12), 1379–1384.
Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*(1), 117–140.
Schneider, P., Dubé, R. V., & Hayward, D. (2005). *The Edmonton Narrative Norms Instrument*. Retrieved November 17, 2015, from http://www.rehabresearch.ualberta.ca/enni
Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language Speech and Hearing Services in Schools*, *41*(4), 459–473.
Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology*, *30*(4), 224–238.
Serratrice, L. (2007). Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers. *Journal of Pragmatics*, *39*(6), 1058–1087.
Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, *27*(6), 960–974.
Skarabela, B., Allen, S. E. M., & Scott-Phillips, T. C. (2013). Joint attention helps explain why children omit new referents. *Journal of Pragmatics*, *56*, 5–14.
Soodla, P. (2011). *Picture-elicited narratives of Estonian children at the kindergarten-school transition as a measure of language competence* (PhD Thesis). University of Tartu, Tartu.
Soodla, P., & Kikas, E. (2010). Macrostructure in the narratives of Estonian children with typical development and language impairment. *Journal of Speech Language and Hearing Research*, *53*(5), 1321–1333.
Statistics Sweden. (2017a). *Befolkningens utbildning 2016* [Educational attainment of the population 2016] (Statistiska meddelanden No. UF 37 SM 1701).
Statistics Sweden. (2017b). *Statistikdatabasen*. Retrieved March 22, 2018 from www.statistikdatabasen.scb.se.
Stein, N. L., & Albro, E. R. (1997). Building complexity and coherence: Children's use of goal-structured knowledge in telling stories. In M. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 5–44). Hillsdale, NJ: Lawrence Erlbaum.
Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Norwood, NJ: Ablex.
Stein, N. L., & Policastro, M. (1984). The concept of a story: A comparison between children's and teacher's viewpoints. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 113–155). Hillsdale, NJ: Lawrence Erlbaum.
Strömqvist, S., & Verhoeven, L. T. (Eds.). (2004). *Relating events in narrative. Vol. 2, Typological and contextual perspectives*. Hillsdale, NJ: Lawrence Erlbaum.
Sundqvist, P. (2009). *Extramural English matters: out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary* (PhD Thesis). Karlstad University, Karlstad.
Tallberg, I.-M. (2005). The Boston Naming Test in Swedish: Normative data. *Brain and Language*, *94*(1), 19–31.
Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Topaj, N. (2010). Topical referential expressions in narratives of Russian-German bilingual children. In M. Chini (Ed.), *Topic, information structure and language acquisition* (pp. 59–72). Milano: Franco Angeli.

Topbaş, S., & Kaçar, D. (2013). A comparative study: The implementation of a Turkish Nonword Repetition Test and Quasi-Universal Nonword Repetition Test with monolingual Turkish children with SLI. Presented at the COST Action IS0804 meeting, February 13th-15th 2013, Lisbon.

Topbaş, S., Kaçar-Kütükçü, D., & Kopkalli-Yavuz, H. (2014). Performance of children on the Turkish Nonword Repetition Test: Effect of word similarity, word length, and scoring. *Clinical Linguistics & Phonetics*, *28*(7–8), 602–616.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*(3), 249–275.

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goal/plans: A conceptual basis for narrating "Frog where are you?" In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: a crosslinguistic developmental study* (pp. 85–106). Hillsdale, NJ: Lawrence Erlbaum.

Trabasso, T., Stein, N. L., Rodkin, P. C., Munger, M. P., & Baughn, C. R. (1992). Knowledge of goals and plans in the on-line narration of events. *Cognitive Development*, *7*(2), 133–170.

Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, *37*(1), 195–216.

Turell, M. T., & Moyer, M. G. (2008). Transcription. In L. Wei & M. G. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 192–213). Malden, MA: Blackwell.

Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language Speech and Hearing Services in Schools*, *38*(3), 225–236.

UNESCO Institute for Statistics. (2012). *International standard classification of education: ISCED 2011*. Montreal, Quebec: UNESCO Institute for Statistics. Retrieved from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

Verhoeven, L., & Strömqvist, S. (Eds.). (2001). *Narrative development in a multilingual context*. Amsterdam: John Benjamins.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*(1), 65–83.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234.

Viberg, Å. (2001). Age-related and L2-related features in bilingual narrative development in Sweden. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative Development in a Multilingual Context* (pp. 87–128). Amsterdam: John Benjamins.

Viberg, Å. (2004). Lexikal utveckling i ett andraspråk [Lexical development in a second language]. In K. Hyltenstam & I. Lindberg (Eds.), *Svenska som andraspråk - i forskning, undervisning och samhälle* (pp. 197–220). Lund: Studentlitteratur.

Waldmann, C. (2008). *Input och Output: Ordföljd i svenska barns huvudsatser och bisatser* [Input and Output: Word order in Swedish children's main and subordinate clauses]. (Lundastudier i nordisk språkvetenskap A65. PhD Thesis). Lund University, Lund.

Warden, D. A. (1976). The influence of context on children's use of identifying expression and references. *British Journal of Psychology*, *67*(1), 101–112.
Westby, C. E. (1984). Development of narrative abilities. In G. P. Wallach & K. G. Butler (Eds.), *Language learning disabilities in school-age children* (pp. 103–127). Baltimore: Williams & Williams.
Westby, C. E. (2012). Assessing and remediating text comprehension problems. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities*. (3rd edition, pp. 163–225). Boston, MA: Pearson.
Westlin, M., & Ytterdahl, O. (2007). *Strategier vid bildbenämning: En explorativ studie av 3-4 åringars ordförråd* [Strategies in picture naming: An explorative study of vocabulary in 3-4-year-olds] (Unpublished MA thesis). Karolinska Institutet, Stockholm.
Whitely, C., & Colozzo, P. (2013). Who's who? Memory updating and character reference in children's narratives. *Journal of Speech Language and Hearing Research*, *56*(5), 1625–1636.
Widaman, K. F. (2006). Missing data: What to do with or without them. In K. McCartney, M. R. Burchinal, & K. L. Bub (Eds.), *Best practices in quantitative methods for developmentalists*. (Vol. 3, pp. 42–64). Boston, MA: Blackwell.
Wigglesworth, G. (1990). Children's narrative acquisition: a study of some aspects of reference and anaphora. *First Language*, *10*(29), 105–125.
Wittek, A., & Tomasello, M. (2005). Young children's sensitivity to listener knowledge and perceptual context in choosing referring expressions. *Applied Psycholinguistics*, *26*(4), 541–558.
Wong, A. M.-Y., & Johnston, J. R. (2004). The development of discourse referencing in Cantonese-speaking children. *Journal of Child Language*, *31*(3), 633–660.

ACTA UNIVERSITATIS UPSALIENSIS

Studia Linguistica Upsaliensia

Editors: Michael Dunn and Joakim Nivre

1. *Jörg Tiedemann*, Recycling translations. Extraction of lexical data from parallel corpora and their application in natural language processing. 2003.
2. *Agnes Edling*, Abstraction and authority in textbooks. The textual paths towards specialized language. 2006.
3. *Åsa af Geijerstam*, Att skriva i naturorienterande ämnen i skolan. 2006.
4. *Gustav Öquist,* Evaluating Readability on Mobile Devices. 2006.
5. *Jenny Wiksten Folkeryd,* Writing with an Attitude. Appraisal and student texts in the school subject of Swedish. 2006.
6. *Ingrid Björk,* Relativizing linguistic relativity. Investigating underlying assumptions about language in the neo-Whorfian literature. 2008.
7. *Joakim Nivre, Mats Dahllöf and Beáta Megyesi,* Resourceful Language Technology. Festschrift in Honor of Anna Sågvall Hein. 2008.
8. *Anju Saxena & Åke Viberg*, Multilingualism. Proceedings of the 23rd Scandinavian Conference of Linguistics. 2009.
9. *Markus Saers*, Translation as Linear Transduction. Models and Algorithms for Efficient Learning in Statistical Machine Translation. 2011.
10. *Ulrika Serrander*, Bilingual lexical processing in single word production. Swedish learners of Spanish and the effects of L2 immersion. 2011.
11. *Mattias Nilsson*, Computational Models of Eye Movements in Reading : A Data-Driven Approach to the Eye-Mind Link. 2012.
12. *Luying Wang,* Second Language Acquisition of Mandarin Aspect Markers by Native Swedish Adults. 2012.
13. *Farideh Okati*, The Vowel Systems of Five Iranian Balochi Dialects. 2012.
14. *Oscar Täckström,* Predicting Linguistic Structure with Incomplete and Cross-Lingual Supervision. 2013.
15. *Christian Hardmeier,* Discourse in Statistical Machine Translation. 2014.
16. *Mojgan Seraji,* Morphosyntactic Corpora and Tools for Persian. 2015.
17. *Eva Pettersson,* Spelling Normalisation and Linguistic Analysis of Historical Text for Information Extraction. 2016.
18. *Marie Dubremetz,* Detecting Rhetorical Figures Based on Repetition of Words: Chiasmus, Epanaphora, Epiphora. 2017.
19. *Josefin Lindgren,* Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6. 2018.