

Acquisition of narrative macrostructure: A comprehensive overview of results from the Multilingual Assessment Instrument for Narratives

Josefin Lindgren*

Uppsala University, Sweden

Freideriki Tselekidou

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Germany

Natalia Gagarina

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Germany

In this paper, we give a comprehensive overview of the results from studies that have used the Multilingual Assessment Instrument for Narratives (MAIN) to investigate comprehension and production of narrative macrostructure (story structure) to date. We show the wide range of research in which MAIN has been used through summaries of core results from studies that investigated age effects, and studies that compared monolinguals with bilinguals, bilinguals' two languages, and typically-developing (TD) children with children with developmental language disorder (DLD). Results from studies including factors that influence bilinguals' narrative skills (e.g., language skills, language input) are also covered, as are those that deal with methodological aspects and more specifically, task effects, i.e., how the choice of elicitation mode (telling; retelling; model story) and story (Cat/Dog; Baby Birds/Baby Goats) influence story structure and story comprehension. As concluding remarks, we summarize the state-of-the-art of narrative research using MAIN and outline possible directions for future studies.

* Corresponding author: josefin.lindgren@nordiska.uu.se

1 Introduction

The Multilingual Assessment Instrument for Narratives (LITMUS MAIN, hereafter MAIN; Gagarina et al., 2012; Gagarina, Klop et al., 2019) was developed within the framework of COST Action IS0804 *Language Impairment Testing in Multilingual Settings* with the aim to assess narrative production and comprehension skills of children from 3 to 10 years old. Later, it has expanded to older children, adolescents, and adults. MAIN was originally published in 2012 (with a revised version in 2019). Currently, it is available in over 90 languages and during the past 10 years, a large number of published studies have used it to investigate different aspects of children's narratives. These studies cover a variety of languages and language combinations. Three special journal issues have been published, in *Applied Psycholinguistics* on *Narrative abilities in bilingual children* (Gagarina et al., 2016), in *First Language* focusing on *Children's acquisition of referentiality in narratives* (Gagarina & Bohnacker, 2022a) and in *Linguistic Approaches to Bilingualism* on *Storytelling in bilingual children* (Gagarina & Bohnacker, 2022b). Additionally, a recent book volume focused on narrative comprehension using MAIN (Bohnacker & Gagarina, 2020). However, so far, no comprehensive overview of the results from studies using MAIN has been published.

In this paper, we therefore summarize the findings of the research using MAIN, focusing on the scores on *story comprehension* and *story structure* (narrative macrostructure), two commonly used measures that are included as standardized measures in the instrument. An overview of all reviewed studies is given in Table 1 in the Appendix.¹ We start by giving an overview of MAIN and these two narrative measures (Section 2). Next, studies investigating age effects and development are described (Section 3), after which we focus on studies of bilingual children, including those who investigated different factors influencing bilinguals' narrative skills (Section 4), and on studies comparing typically-developing (TD) children with children with developmental language disorder (DLD) (Section 5). In the following section (Section 6), we describe results from studies investigating methodological aspects and more specifically, task effects, i.e., how results are influenced by the choice of elicitation mode and story. Finally, in our concluding remarks, we summarize what these studies tell us about children's narrative abilities and outline possible future directions for narrative research (Section 7).

2 The Multilingual Assessment Instrument for Narratives: an overview

Based on the need to create a satisfactory and theoretically-grounded instrument for the assessment of children's narrative skills across different languages, MAIN was developed by an interdisciplinary group of researchers. It consists of four parallel picture-based stories (Cat,

¹ In our review, we include only those published studies that analysed narratives elicited with MAIN and report results for the story structure score (or a close equivalent that included the same types of macrostructural components) and/or the standardized measure of story comprehension (see Section 2). We have done our best to find all such studies, but there may be additional studies of which we are not aware. Studies of MAIN narratives that only investigate other measures (e.g., reference or other aspects of microstructure) were not included.

Dog, Baby Birds, Baby Goats), each consisting of six pictures accompanied by story scripts. The four stories are equivalent in terms of their linguistic and cognitive demands, and are controlled for their cultural appropriateness. Thus, they allow testing both languages of bilinguals and make it possible to draw parallels between children from different linguistic and cultural backgrounds. Children's narrative abilities can be assessed via three elicitation procedures, telling, retelling, and model story, and the resulting narratives can be analyzed on both microstructural and macrostructural levels. The microstructure focuses on aspects that are connected to language specific characteristics, such as the number and complexity of words and sentences, and lexical diversity, while macrostructure evaluates the higher-order narrative structure, which is thought to express universal structures (cognitive schemata) and can therefore be considered language-independent.

The theoretical fundament of the macrostructure is the multidimensional theory of narrative organization. The core of this theory operationalizes narratives macrostructure in production and comprehension as several layers on the qualitative/quantitative and factual/inferred dimensions. The macrostructure in MAIN consists of single elements, such as Goals (G), internal states (IS), Attempts (A), and Outcomes (O), which are organized in episodes. With respect to the dimension of quantity/quality, the number (or sum) of these produced components is the *story structure*, the quantitative measure of the macrostructure. The MAIN story structure assessment consists of awarding points for the production of a setting (time + place, Max=2) and for IS as initiating event, goal, attempt, outcome, and IS as reaction in each of the three episodes (Max=15 points, 3x5 components). The maximum story structure score in all MAIN-stories is thus 17, if all components specified in the model are verbalized by the narrator.

Story complexity measures another dimension of a narrative. It reflects the qualitative aspect of narrative macrostructure by examining the ability to combine the core episodic components (e.g., A+O, G+A, G+O, G+A+O) in order to form episodes with different levels of complexity, with the most complex episode being a so-called complete episode, a goal-attempt-outcome sequence (GAO). Story complexity can be operationalized in different ways, e.g., by counting the number of GAO-sequences (max=3) produced by a child or by analyzing the proportions of different types of sequences found in narratives produced by a group of children. In previous studies, story complexity has indeed been operationalized in a number of ways that are not fully comparable. For this reason, in the present paper, we only report results from studies investigating story structure.

Another dimension of the multidimensional theory is factual vs inferred components. Goals and internal states are the two inferred components that are present in each of the three episodes in MAIN. Factual components, such as attempts and outcomes, are visualized directly in the pictures, and they are therefore easier to recognize, produce and understand. The dimension factual vs. inferred is reflected only in the production, since the comprehension questions target only inferred types of components (G and IS).

Story comprehension is assessed via 10 questions,² which are asked after the production of the story (or after the child has listened to a model story), and target inferred components of narrative macrostructure. Answering the questions correctly requires the child to use Theory of Mind to understand what the characters want and feel at specific points in the story. Three questions tap into the child's understanding of the three goals (one in each episode), six target the character's internal states and the child's ability to express the character's reason for experiencing these internal states, and one question assess the child's understanding of the whole plotline. Thus, the comprehension questions allow the researcher to draw conclusions with respect to different types of inferred information.

The quality of MAIN as a testing instrument was assessed in a study by Lautenschläger et al. (2021), who performed a psychometric evaluation analysing its objectivity, reliability, and validity. In their study, focus was on the story structure score and on the Baby Birds and Baby Goats stories. Generally, the instrument performed satisfactorily, although with some differences between different measures. Using some additional guidelines/criteria next to the MAIN scoring protocol, the interrater agreement (i.e. objectivity) was very high (98.13%). The test-retest reliability showed an almost perfect correlation ($r=.978$) between two testing points with around two weeks in between, when the same story was used twice. The children received significantly higher scores on the second testing point, showing a training effect. There was a lower reliability when two different stories were used: the correlation between the scores when the children told both stories was substantially lower, but still high ($r=.767$), and there was no significant difference between Baby Birds and Baby Goats. To investigate the validity of MAIN, the story structure scores were correlated with scores from a test of expressive vocabulary. The result showed a strong positive correlation, that was nevertheless not perfect ($r=.648$), between story structure and expressive vocabulary, indicating that although the story structure score is closely linked to expressive vocabulary (see also the results presented in Section 4.3 below), it does measure something in addition to the child's vocabulary skills. The authors conclude that further investigation into the validity of MAIN is needed, that it is important that scorers of MAIN receive extensive training to have satisfactorily objectivity in the scoring, and that researchers need to be aware of the fact that that when two different stories are used (e.g., in a bilingual child's two languages), there will be some difference in the scores.

3 Age effects and development

The majority of the studies using MAIN have focused on age 4 to 7, i.e., children attending preschool and/or the first grades of primary school, depending on the school system of the country (e.g., Altman et al., 2016; Boerma et al., 2016; Bohnacker et al., 2020; Bohnacker & Lindgren, 2021; Fichman et al., 2022; Haddad, 2022; Kapalková et al., 2016; Kunnari et al., 2016; Lindgren, 2018; Öztekin, 2019; Peristeri et al., 2020; Roch et al., 2016; Wehmeier, 2020), with fewer studies including children aged 8 or above (e.g., Fiani et al., 2020, 2022; Gagarina,

² In a few early studies, the participants were only asked 9 questions, as the final question was added a bit later in the process of developing MAIN.

2016; Košutar et al., 2022; Lindgren, 2022; Peristeri et al., 2020; Tribushinina et al., 2022; Tsimpli et al., 2016; Yang et al., 2023). Relatively many previous studies using MAIN analysed data from participants of different ages without including age as a variable in the study (e.g., Altman et al., 2016; Blom & Boerma, 2020; Boerma et al., 2016; Fichman et al., 2022; Kunnari & Välimaa, 2020). To our knowledge, only two studies have focused solely on adults. Gagarina, Bohnacker, et al. (2019) analyzed story structure and story complexity in German-, Russian-, and Swedish-speaking adults (N=69) and Antonijevic et al. (2022) investigated story structure and story comprehension in Irish-English bilinguals (N=30).

Studies investigating age effects on *story comprehension* have mostly found clear age effects for age 3/4–6/7, both in monolinguals (e.g., Bohnacker & Lindgren, 2021; Lindgren, 2019, 2022) and in bilinguals (e.g., Bohnacker, 2016; Bohnacker et al., 2020; Gagarina et al., 2020; Haddad, 2022; Lindgren & Bohnacker, 2020). The studies by Roch and Hržica (2020) and Blom and Boerma (2016) form exceptions here. Roch and Hržica (2020) did not find a significant correlation between age and story comprehension in Croatian-Italian bilingual 5–7-year-olds (N=30), and in their longitudinal study of Dutch monolingual children with TD (N=45) and with DLD (N=84), Blom and Boerma (2016) found no development from age 5–6 to age 6–7. There are some indications that story comprehension is close to ceiling already at age 5–6 (Bohnacker & Lindgren, 2021; Lindgren, 2019, 2022), which may explain these findings. However, Fiani et al. (2020), in their study of Lebanese Arabic-French bilinguals (N=48) found a significant effect of age from age 4 to 9, and likewise did Peristeri et al. (2020) from age 6 to 9 in a study of Albanian-Greek bilinguals and Greek monolinguals. In some studies, the age effect differed between groups or languages. For example, Wehmeier (2020), in her study of German-speaking monolinguals (N=199) and bilinguals (N=66) aged 4;6–5;11, found an age effect only in the monolingual group; this may have been due to the smaller bilingual sample. Rodina (2017), in her study of Russian-Norwegian bilinguals aged 4–6 (N=16), found an age effect on story comprehension in Norwegian, but not in Russian.

With respect to *story structure*, development have also generally been found between ages 3–4 and 6–7, for both monolinguals (Lindgren, 2019, 2022) and bilinguals (Bohnacker, 2016; Bohnacker et al., 2022; Fiani et al., 2022; Gagarina, 2016; Haddad, 2022; Lindgren & Bohnacker, 2022; Roch et al., 2016). The study by Blom and Boerma (2016) cited above is again an exception, with no significant development in story structure from age 5–6 to 6–7.

Studies including older children show more mixed results. For example, in a study of Croatian-speaking monolinguals (N=89), Košutar et al. (2022) found a significant difference between ages 6 and 8, whereas Gagarina (2016), in a study of Russian-German bilinguals found no significant difference between children in Grade 1 (aged 6;5–7;5) and Grade 3 (aged 7;11–10;6). Similarly, Fiani et al (2022) found no significant difference between Lebanese Arabic-French bilinguals aged 6–7 and 8–9, and Yang et al. (2023), in their study of Kam-Mandarin Chinese bilinguals aged 5 to 9 (N=55), found no effect of age. In a longitudinal study from age 4 to 7 of narratives elicited with the Baby Birds/Baby Goats stories from Swedish-speaking monolinguals (N=17), Lindgren (2019) found no difference in the story structure between age 5;10 and 7;4, whereas a subsequent study of the same children's Cat/Dog narratives from age 4 to 9 (Lindgren, 2022) found a significant development from age 7 to age 9. It is thus possible

that children's story structure reaches a plateau before it develops further. Again, some studies found differences between groups or languages. Tribushinina et al. (2022) found an age effect in the home language Indonesian of Indonesian-Dutch bilinguals aged 5–12 (N=32), but not in the societal language Dutch. More research is thus needed on children above the age of 7–8 to be able to draw firm conclusions as to how narrative skills develop further and at what age these skills reach the level of adults.

4 Bilinguals' narrative abilities

The majority of the published studies using MAIN have investigated narrative abilities of bilingual children. Some studies compared them to monolinguals (e.g., Blom & Boerma, 2020; Boerma et al., 2016; Bohnacker & Lindgren, 2021; Kunnari et al., 2016; Kunnari & Välimaa, 2020; Peristeri et al., 2020; Rodina, 2017; Tsimpli et al., 2016), others compared different bilingual groups (e.g., Blom & Boerma, 2020), or bilinguals' two languages (e.g., Altman et al., 2016; Bohnacker, 2016; Bohnacker et al., 2020, 2022; Bohnacker & Lindgren, 2021; Fiani et al., 2020, 2022; Fichman et al., 2022; Gagarina, 2016; Kapalková et al., 2016; Kunnari et al., 2016; Kunnari & Välimaa, 2020; Lindgren & Bohnacker, 2020, 2022; Öztekin, 2019; Roch et al., 2016; Roch & Hržica, 2020; Rodina, 2017; Tribushinina et al., 2022). One study combined all three types of comparisons (Lindgren, 2018). Several studies using MAIN have also investigated the effect of background factors or general language skills on bilinguals' narrative skills. Here, we summarize results from studies comparing bilinguals and monolinguals or different bilingual groups (Section 4.1), comparing bilinguals' two languages (Section 4.2) and those investigating factors influencing bilinguals' narrative skills (Section 4.3).

4.1 *Bilinguals vs monolinguals and comparisons of different bilingual groups*

The majority of studies comparing monolinguals and bilinguals in *story comprehension* have found that the groups do not differ significantly from each other (Blom & Boerma, 2020; Boerma et al., 2016; Bohnacker & Lindgren, 2021; Kunnari & Välimaa, 2020; Rodina, 2017). For example, Boerma et al. (2016) found that Dutch-speaking monolinguals and bilinguals (N=132) performed similarly on story comprehension after both telling and model story. One study forms an exception: in the study by Peristeri et al (2020), Albanian-Greek bilingual 6–8-year-olds were found to perform significantly better on story comprehension in the retelling mode than their monolingual Greek-speaking peers.³ The study by Lindgren (2018), which compared Swedish monolinguals (N=72), German-Swedish bilinguals (N=46), and Turkish-Swedish bilinguals (N=48), found that the latter group performed significantly lower than the other two on comprehension of the Baby Birds/Baby Goats, but not Cat/Dog (both used in the telling mode). This indicates that the performance of different groups may be influenced by the type of stimulus material. Lindgren (2018) also found a significant difference between the two bilingual groups in their home languages, where the German-Swedish group had significantly higher comprehension scores in German than the Turkish-Swedish group in Turkish. In their

³ The home language Albanian was not investigated.

longitudinal study with three timepoints (T1 at age 5–6) of comprehension after model story and telling, Blom and Boerma (2020) compared story comprehension for different bilingual groups, Turkish-Dutch (N=31) and Tarifit-Dutch (N=38). They only found a significant difference for story comprehension after telling at T2 (age 6–7), with higher scores in the Turkish-Dutch group.

For *story structure*, results are a bit more mixed. A number of studies found that bilinguals and monolinguals perform similarly also for this measure (Boerma et al., 2016; Haman et al., 2017; Kunnari et al., 2016). Tsimpli et al. (2016), using a modified scale of story structure where points were only awarded for the production of goals, attempts, and outcomes (Max=9 points), found that Greek-speaking bilinguals performed significantly higher than monolinguals. Rodina (2017) found that Russian-Norwegian bilinguals aged 4 to 6 (N=16) growing up in Norway performed similarly to Norwegian monolinguals, but significantly lower than Russian monolinguals. In the study described above, Lindgren (2018) found that the Turkish-Swedish bilinguals performed significantly lower on story structure than both Swedish monolinguals and German-Swedish bilinguals, with no difference between the two latter groups. When comparing story structure in the home language, the German-Swedish group performed significantly higher for narratives elicited with Cat/Dog, but for Baby Birds/Baby Goats there was no significant difference. The reason for this difference is not clear, and further studies that investigate narrative production by monolinguals and bilinguals for the different MAIN stories, while also taking factors such as general language skills into account (see Section 4.3), are still needed.

4.2 Bilinguals' two languages

A relatively large number of studies have compared bilinguals' performances in the two languages and the majority came to the same conclusion, namely that bilinguals perform similarly in the languages. This has been shown for both *story comprehension* (Bohnacker, 2016; Bohnacker et al., 2020; Fiani et al., 2020; Kapalková et al., 2016; Kunnari & Välimaa, 2020; Lindgren & Bohnacker, 2020; Rodina, 2017) and *story structure* (Altman et al., 2016; Bohnacker, 2016; Bohnacker et al., 2022; Fiani et al., 2022; Fichman et al., 2022; Kunnari et al., 2016; Lindgren & Bohnacker, 2022; Rodina, 2017). However, some studies did find differences between the languages. With one exception, these studies found higher scores in the children's first language (L1), irrespective of which language was the societal language. For example, Roch and Hržica (2020) found for story comprehension that Croatian-Italian bilingual 5–7-year-olds growing up in Croatia performed significantly higher in L1 Croatian than in L2 Italian. Similarly, Tribushinina et al. (2022) found higher story structure scores in L1 Indonesian than in L2 Dutch for Indonesian-Dutch bilinguals aged 5–12 growing up in the Netherlands, and Kapalková et al. (2016) found higher story structure scores in L1 Slovak than in L2 English in bilingual 5–6-year-olds growing up in Slovakia. Interestingly, Roch et al. (2016) only found a difference between L1 Italian and L2 English in the younger group (age 5–6), whereas the older group (aged 6–7) showed similar performance in the two languages. The pattern was the same for both story comprehension and story structure. In all cases, these results could be explained by the bilinguals' possibly having lower language proficiency in the L2

compared with the L1. The only exception to the pattern of higher scores in the L1 was found in a study by Lindgren and Bohnacker (2022) of German-Swedish 4–6-year-olds (N=46); these children had significantly higher score in the societal language Swedish than in the home language German. However, the majority of these children were simultaneous bilinguals, speaking both Swedish and German at home from birth, which may explain the higher scores in the societal language, as they may well be more proficient in the this language. These results are thus linked to the studies in the following section, which deals with factors influencing bilinguals' narrative skills.

4.3 *Factors influencing bilinguals' narrative skills*

A number of studies using MAIN have investigated factors influencing bilinguals' narrative skills, including measures of general language skills such as vocabulary and/or grammar (Bohnacker et al., 2020, 2022; Fiani et al., 2020, 2022; Lindgren & Bohnacker, 2020, 2022; Roch & Hržica, 2020; Tsimpli et al., 2016; Yang et al., 2023), measures connected to the amount of input received in the languages, the child's Length of Exposure (LoE) or Age of Onset (AoO) (Blom & Boerma, 2020; Bohnacker et al., 2022; Haman et al., 2017; Lindgren & Bohnacker, 2022; Roch & Hržica, 2020; Tribushinina et al., 2022) as well as language dominance or type of bilingualism (Fiani et al., 2020; Fichman et al., 2022; Gagarina, 2016).

With respect to measures of bilinguals' *language skills*, in a number of studies, the child's expressive vocabulary (measured by a score on a vocabulary test) has been found to significantly influence bilinguals' narrative skills; children's scoring higher on expressive vocabulary have been found to have higher scores on both story comprehension (Bohnacker et al., 2020; Fiani et al., 2020) and story structure (Bohnacker et al., 2022; Fiani et al., 2022; Yang et al., 2023) in both languages.⁴ However, there are indications that, at least for some bilingual groups, the effect of vocabulary may differ between the languages. Two studies of the same German-Swedish bilinguals, one of story comprehension (Lindgren & Bohnacker, 2020) and one of story structure (Lindgren & Bohnacker, 2022), found a significant effect of expressive vocabulary in the home language German but not in the societal language Swedish. The authors propose that this may be linked to the children's overall higher proficiency in the societal language, compared to the home language. Additionally, the only study investigating receptive vocabulary also found different patterns in the two languages, with a significant correlation between receptive vocabulary and story comprehension in L1 Croatian, but not in L2 Italian (Roch & Hržica, 2020). Roch & Hržica (2020) also investigated the correlation between grammatical competence and story comprehension, which was found to be significant in both languages.

Regarding the effects of *language input/exposure* (including measures such as LoE and AoO), results are somewhat mixed and seem to depend both on the bilingual groups investigated and on the measure used. A number of studies have used parental estimates of the children's daily input in the two languages or measures of input at home, finding no effect on

⁴ Yang et al. (2023) only investigated the home language Kam of Kam-Mandarin Chinese bilinguals. In addition to expressive vocabulary, they also found a significant effect of grammar (morphosyntactic ability measured via a sentence repetition task) on the children's story structure score.

or relationship with story comprehension (Blom & Boerma, 2020; Bohnacker et al., 2020; Lindgren & Bohnacker, 2020; Roch & Hržica, 2020)⁵ or story structure (Bohnacker et al., 2022; Lindgren & Bohnacker, 2022). Tribushinina et al. (2022) analyzed current amount and richness of input in relationship to the story structure of Indonesian-Dutch bilinguals, but found no significant effect in either language. However, Haman et al. (2017) found that both higher exposure to L1 Polish and L2 English led to higher story structure scores in the L1, and Peristeri et al. (2020) found an effect of the amount of exposure to Greek before the age of schooling (age 6) on Albanian-Greek 6–8-year-old bilinguals' story comprehension. Roch and Hržica (2020) found no correlation between AoO and story comprehension in Croatian-Italian bilinguals' both languages, and Bohnacker et al. (2022) found no effect of LoE on Turkish-Swedish bilinguals' story structure in both languages. Tribushinina et al. (2022) found no effect of LoE to the majority language Dutch on the Indonesian-Dutch bilinguals' story structure in Dutch.

Finally, two studies have analysed narrative skills in relation to *language dominance*, one on story comprehension in Lebanese Arabic-French bilinguals (Fiani et al., 2020) and one on story structure in Russian-Hebrew bilinguals (Fichman et al., 2022). None of the studies found a significant effect of language dominance in either language, possibly due to the relatively small number of children in each dominance group. Differences between simultaneous and successive bilinguals have only been investigated in one previous study, the study by Gagarina (2016) on Russian-German bilinguals. Simultaneous bilinguals were found to have an advantage on story structure over successive bilingual, but only in the majority language German; no difference was found in the home language Russian. It can thus be concluded that while there seem to be a clear link between bilinguals' language skills and their story structure and story comprehension, factors such as input, LoE, AoO and language dominance need to be investigated further in future studies.

5 Typically-developing children and children with developmental language disorder

A number of studies have used MAIN to compare children with TD and children diagnosed with DLD or children who were identified to be at risk for DLD. Some of the studies investigated monolingual children (Blom & Boerma, 2016; Kuvač Kraljević et al., 2020; Pham et al., 2019; Sheng et al., 2020), others focused only on bilinguals (Altman et al., 2016), whereas yet others included both monolinguals and bilinguals (Boerma et al., 2016; Peristeri et al., 2020; Tsimpli et al., 2016). The specific ages investigated differ between studies, but the majority falls within the range of age 5 to 8.

Only a few studies have investigated *story comprehension* in children with DLD, and results are somewhat mixed. In their longitudinal study of Dutch monolinguals, Blom and Boerma (2016) found a difference between the TD and DLD groups at age 5–6, but not at age

⁵ In their longitudinal study from age 5 to 7 of Turkish-Dutch and Tarifit-Dutch children, Blom and Boerma (2020) only found a significant correlation between home language richness, a measure of the child's input "from family friends and peers, as well as during reading activities, watching television/movies, and oral storytelling" (p. 209), and story comprehension in the home language on the Tarifit-Dutch group at T1 (age 5–6).

6–7, possibly due to ceiling effects. The result for age 5–6 was replicated in a study that also included bilingual Dutch-speaking children with and without DLD (Boerma et al., 2016). Peristeri et al. (2020), in a study of children aged 6–8 (N=120), found a difference between TD and DLD for Greek-speaking monolinguals, but not for Albanian-Greek bilinguals. Further studies of story comprehension in TD-children and children with DLD are thus needed before it is possible to draw any conclusions.

With respect to *story structure*, a larger number of studies have been carried out and results are a bit more uniform. The studies by Blom and Boerma (2016) and Boerma et al. (2016) of Dutch-speaking children, as well as the studies by Kuvač Kraljević et al. (2020) of Croatian-speaking monolinguals and Pham et al. (2019) of Vietnamese-speaking monolinguals, found significantly higher scores in TD-children. The study by Sheng et al. (2020) of Mandarin-speaking children with TD (N=21) and those at risk for DLD (N=21) found a difference between the two groups on story structure in narratives elicited in the telling mode, but no difference between the groups in the retelling mode. Two studies who did not use the story structure score, but instead analyzed a score for story complexity for two narratives combined (Tsimpli et al., 2016) or counted only goals, attempts and outcomes in the narratives (Altman et al., 2016), respectively, did not find any difference between children with TD and DLD. Since these measures differ both from the story structure score and from each other, it is difficult to compare the studies and for this reason, the picture of how DLD influences children’s story structure in MAIN-narrative is still not conclusive.

6 Task effects

6.1 Differences between the stories

The four MAIN-stories, *Cat*, *Dog*, *Baby Birds* and *Baby Goats* were created to be parallel in their macrostructure, and the same number and types of macrostructural components are depicted in or can be inferred from the pictures and are included in the story scripts. However, there are some differences between, on one hand, *Cat/Dog* and, on the other, *Baby Birds/Baby Goats*. In *Cat/Dog*, multiple events that belong to different episodes take place simultaneously, whereas events are organized in a more linear fashion in *Baby Birds/Baby Goats*. *Cat* and *Dog* also contain three characters, where one is human (a boy), whereas *Baby Birds* and *Baby Goats* have five characters, who are all animals. Additionally, in the *Baby Birds* story, the two baby birds together function as one character, whereas the two baby goats in the *Baby Goats* story are separate entities, who are part of different plotlines. These differences may influence the structure of the narratives that children tell to these stories. A number of studies have investigated differences between the stories in story comprehension or story structure, either between the two pairs of stories (i.e., *Cat/Dog* vs. *Baby Birds/Baby Goats*) or between two specific stories (e.g., Bohnacker et al., 2022; Fichman et al., 2022; Gagarina, Bohnacker, et al., 2019; Kavar et al., 2023; Lindgren & Bohnacker, 2020). Here, we summarize the findings from these studies. The studies reported here used the telling mode unless stated otherwise. Studies employing different stories in different elicitation modes, e.g., *Cat/Dog* in retelling and *Baby Birds/Baby Goats* in telling, are discussed in Section 6.2.

For *story comprehension*, a number of studies have found a significant difference between Cat/Dog and Baby Birds/Baby Goats, when the same children told one story from each pair (i.e., a within-subjects design), with higher scores on the former than on the latter (Bohnacker et al., 2020; Bohnacker & Lindgren, 2021; Kawar et al., 2023; Lindgren & Bohnacker, 2020).⁶ The difference has been found both for monolinguals and in bilinguals' two languages. For example, in their study of 100 Turkish-Swedish bilinguals aged 4 to 7, Bohnacker et al. (2020) found significantly higher scores on story comprehension of Cat/Dog than Baby Birds/Baby Goats in both languages. In all these studies, all children told Cat/Dog first, and the significant difference could thus be caused by an order effect, i.e., that the children were experiencing fatigue while answering the questions to Baby Birds/Baby Goats at the end of the testing session.⁷ A number of studies have compared performance on story comprehension between the two stories within a pair, where different children told different stories (i.e., a between-subjects design). Here, no significant differences have been found between Cat and Dog (Bohnacker et al., 2020; Bohnacker & Lindgren, 2021; Lindgren, 2018, 2022), whereas scores have been found to be significantly higher on Baby Goats than on Baby Birds, for monolinguals (Bohnacker & Lindgren, 2021; Lindgren, 2019) and for bilinguals, at least in one of the languages (Bohnacker et al., 2020, for Turkish of Turkish-Swedish-bilinguals; Bohnacker & Lindgren, 2021, for English of English-Swedish bilinguals). For story comprehension there are thus indications that, on the one hand, the pairs Cat/Dog and Baby Birds/Baby Goats may not be completely comparable, but also that Baby Goats and Baby Birds may differ.

For *story structure*, studies comparing narratives elicited with Cat/Dog and Baby Birds/Baby Goats from the same children have found no significant differences (Bohnacker et al., 2022; Lindgren, 2018; Lindgren & Bohnacker, 2022), despite the fact that Cat/Dog was always administered first. Studies comparing Cat and Dog have found no significant differences (Lindgren, 2018, 2022; Öztekin, 2019), whereas results are more mixed for Baby Birds and Baby Goats. A number of studies have found that these two stories did not differ significantly either (Fichman et al., 2022, using the retelling mode; Lindgren, 2018; Öztekin, 2019), whereas others have found significantly higher scores on Baby Goats than on Baby Birds (Gagarina, Bohnacker, et al., 2019; Lindgren, 2019). There are thus some indications that, on the one hand, the story pairs Cat/Dog and Baby Birds/Baby Goats may lead to equivalent performance on story structure, but on the other hand, there may be a difference between Baby Birds and Baby Goats. These issues need further investigation.

6.2 Differences between telling, retelling and model story

As described above, MAIN can be administered in three different modes, *telling*, *retelling* and *model story*. A number of published studies have compared the telling and the retelling mode,

⁶ Note that, in Lindgren and Bohnacker (2020), the difference was significant for the 4- and 5-year-olds but not for the 6-year-olds, which was likely due to the high scores on both tasks in this group. In Kawar et al. (2023), the retelling mode was used for both Cat/Dog and Baby Birds/Baby Goats, and all children retold all four stories.

⁷ In Kawar et al. (2023), it is not clear whether the order of Cat/Dog and Baby Birds/Baby Goats was randomized or not.

investigating story comprehension (Kunnari & Välimaa, 2020; Wehmeier, 2020), story structure (Kunnari et al., 2016; Kuvač Kraljević et al., 2020; Sheng et al., 2020) or a combination of the two (Maviş et al., 2016; Otwinowska et al., 2020; Roch et al., 2016; Wehmeier, 2019). Some studies compared story comprehension after model story and after telling (Blom & Boerma, 2020; Gagarina et al., 2020; Maviş et al., 2016).

The studies follow the same general procedure (as described in the MAIN manual), but there are some minor differences between studies in the retelling/model story procedure used. In the studies by Roch et al. (2016) and Kuvač Kraljević et al. (2020), the child listened to a pre-recorded story in headphones, whereas in the other studies, an experimenter read the story scripts.⁸ In most studies, in both modes, only the child was able to see the pictures (as per the standardized procedure); the study by Otwinowska et al. (2020) forms an exception here. In Otwinowska et al. (2020), the pictures were visible to both experimenter and child during the retelling, whereas the telling was done with the pictures visible to the child only. All studies published so far have used the Cat/Dog stories for model story/retelling and the Baby Birds/Baby Goats stories for telling,⁹ which was the intention in the original version of MAIN (Gagarina et al., 2012), but was changed in the revised MAIN (Gagarina, Klop, et al., 2019). Additionally, only the study by Roch et al. (2016) counterbalanced the order of the retelling and telling tasks; in the other studies, all children received the tasks either in the order retelling (or model story) followed by telling (Blom & Boerma, 2020; Gagarina et al., 2020; Maviş et al., 2016, study 1; Sheng et al., 2020; Wehmeier, 2019, 2020) or telling followed by retelling (Kunnari et al., 2016; Kunnari & Välimaa, 2020; Kuvač Kraljević et al., 2020; Maviş et al., 2016, study 2; Otwinowska et al., 2020).

For *story comprehension*, results with respect to the effect of elicitation mode are somewhat mixed. Three studies have found a significant difference between retelling and telling in story comprehension (Otwinowska et al., 2020; Roch et al., 2016; Wehmeier, 2020), with higher scores for the former than the latter. For example, in their study of children aged 5–7 with L1 Italian and L2 English (N=62), Roch et al. (2016) found significantly higher scores in retelling than in telling in both languages. The difference between retelling and telling was large (around 2 points, with max=9 points). However, two studies found no significant differences (Kunnari & Välimaa, 2020; Maviş et al., 2016, study 2). It is notable that the studies which found no significant difference were smaller than those who found an effect, i.e., the former studies may have suffered from a lack of power. Three studies (Blom & Boerma, 2020; Gagarina et al., 2020; Maviş et al., 2016, study 1) investigated comprehension after model story and comprehension after telling, two of them longitudinally. Blom and Boerma (2020) analysed data from Dutch-speaking monolinguals (N=45) and bilinguals (N=69) at three testing points with approximately one year between. They found a significant difference between model story and the telling comprehension at T1 (age 5–6) and T2 (age 6–7), but not at T3 (age 7–8), due to ceiling effects; at T3, scores were above 90% (9 points) in both modes. Similarly, Gagarina

⁸ In two studies (Blom & Boerma, 2020; Wehmeier, 2020), it was not specified who read the story scripts or how the child heard them, but it can be assumed that the more common procedure, where the experimenter reads the story, was used.

⁹ Wehmeier (2019, 2020) used the Cat story for retelling and the Baby Birds for telling.

et al. (2020), in their three-year longitudinal study of Russian-German (N=30) and Turkish-German (N=27) bilinguals, who were either aged 2;10-3;11 or 3;0-4;7 at T1, only found a significant effect of mode in the younger group, possibly due to ceiling effects. The effect of mode was also only found for the comprehension of goals, not ISTs.

With the exception of the study by Otwinowska et al. (2020), which added an extra three points for the production of GAOs to the story structure score, the studies investigating task effects on *story structure* used the same standardized story structure score. The results for story structure are even more mixed compared to those for story comprehension. Roch et al. (2016) found a significant, but relatively small difference in the story structure score, with higher scores in retelling. Similarly, the studies by Otwinowska et al. (2020), Kuvač Kraljević et al. (2020) and Wehmeier (2019) also found significantly higher scores in retelling. However, Maviş et al. (2016, study 2) found no significant difference between the elicitation modes, and Kunnari et al. (2016) found a significant difference for bilinguals in Finnish, but not in Swedish and no significant difference in Finnish monolinguals. Sheng et al. (2020) found no effect of elicitation mode in TD children, but significantly higher scores in retelling than in telling for children who were at risk for DLD.

There are thus some indications that elicitation mode influences both story structure and story comprehension, but results are somewhat mixed, and may be related to sample sized. Due to the fact that the studies have used different stories in the two modes, it is also possible that the differences found between model story/retelling and telling in some previous studies is in fact an effect of differences between the stories (see also Section 6.1). Similarly, in all studies except the study by Roch et al. (2016), the difference could potentially be caused by either a learning effect or an exhaustion effect, depending on which task was administered first. The fact that Roch et al. (2016) as well as studies using different orders (retelling-telling or telling-retelling) have found similar results, namely higher performance in retelling than in telling indicate that the effect is likely not caused by the order, but is a true effect, either of elicitation mode or differences between Cat/Dog and Baby Birds/Baby Goats. Future studies investigating effects of elicitation mode should control for the stories used in the different modes as well as the order of the tasks.

7 Concluding remarks

The aim of this paper was to conduct a comprehensive review of the results from research using MAIN to investigate children's narrative skills. We have summarized findings from studies of comprehension and production of narrative macrostructure (story structure), focusing on how these two narrative measures (the story comprehension and story structure scores) are affected by age, bilingualism and factors influencing bilinguals' narratives (e.g., language skills, input), language, and DLD. We also focused on methodological aspects, more specifically task effects, i.e., the choice of elicitation mode (telling; retelling; model story) and story (Cat/Dog; Baby Birds/Baby Goats). In these concluding remarks, we summarize the state-of-the-art of narrative research using MAIN and outline possible directions for future studies.

With respect to *age*, a clear development has been found during the preschool age (age 3 to 7), but results from those few studies that investigated older children are more mixed. A central issue for future studies is thus to focus on children above age 7 to extend our knowledge of how narrative skills develop further and at what age these skills reach the level of adults.

Most studies using MAIN have investigated *bilingual* children, comparing them to monolinguals, comparing different bilingual groups or comparing bilinguals' two languages. Some have also investigated the effects of various background factors, most notably general language skills. Results indicate that bilinguals often do not differ from monolinguals, and that bilinguals tend to perform similarly in their two languages, but that factors such as the stimuli used and the language skills of the bilinguals in each of the languages play a role. General language skills, such as vocabulary knowledge, have been found to influence bilinguals' comprehension and production of narrative macrostructure. Studies investigating the effect of language exposure/input show mixed results; here results may depend on the characteristics of the studied group as well as on the measure used. It is thus necessary to be cautious with generalizations with respect to how such factors influence bilinguals' narrative skills. Further studies investigating the effects of a wider range of background factors on larger groups of bilinguals are still needed. Additionally, further studies comparing monolinguals and bilinguals speaking various languages are essential in order to fully verify whether the MAIN measures of story structure and story comprehension are indeed independent of the language spoken and the cultural context in which children grow up, as they were designed to be.

A number of studies have compared TD-children and children with *DLD*, but results are mixed, both for comprehension and production of narrative macrostructure, and, additionally, few studies have used the standardized story structure score. For this reason, further studies that use comparable measures and investigate a larger number of languages and language combinations are needed before any firm conclusions can be drawn as to the influence of *DLD* on children's story structure and story comprehension.

Concerning *task effects*, the results from previous studies indicate that researchers need to be cautious when using the different stories. The Cat/Dog and Baby Birds/Baby Goats stories, despite being constructed to be parallel both in their macrostructure and in their comprehension questions, may not be completely comparable. This also has consequences for studies investigating the effects of elicitation modes; future studies comparing for example telling and retelling should control for the specific stories used in the different modes. Since results from some previous studies, including the psychometric evaluation carried out by Lautenschläger et al. (2021), indicate that there may be differences between Baby Goats and Baby Birds, we recommend future studies to use Cat/Dog in situations when total comparability is needed. Additionally, as it is still unknown how the order of the tasks may influence performance, when comparing telling and retelling, it is also necessary to counterbalance the order of the tasks across participants.

From these summaries, it is clear that previous research on MAIN has helped us gain important knowledge about children's narrative skills, especially since the instrument has been used with participants who speak a wide range of languages, come from different cultures and belong to different age groups (see Table 1 in the Appendix), but that a number of issues, which

we have pointed out above, still remain open. In addition to investigating these issues further, what directions could future research using MAIN take? There are of course multiple answers to this question, depending on the specific interest of the researcher as well as the needs of the specific social and cultural contexts in which the study takes place. However, we want to stress a few points that we find especially important.

First, we suggest that future research on MAIN could work towards establishing norms, or at least referential norms, for the acquisition of different narrative skills. Such referential norms must consider the child's chronological age, the Age of Onset of bilingualism, language use and input, factors that would need to be operationalized as scores and be part of the referential norms. The scores on story comprehension and story structure, being the most frequently used measures so far, would be the logical point to start this work. However, establishing norms would require the pooling of resources from a large number of researchers. Therefore, this must be seen as a long-term goal.

Second, another fruitful avenue to go down would be to develop the theoretical model of multidimensional narrative organization further. This endeavour would contribute to a deeper knowledge of narrative skills and their development and form the starting point for new lines of research. To develop the multidimensional model of narrative organization further, a large dataset that contains oral (and possibly also written) narratives from adult speakers of different languages is needed, in addition to the existing datasets from child speakers.

Third, we see a need for further research that includes in-depth analyses of narrative microstructure, including both lexical and grammatical measures. Results from previous studies indicate that narrative skills are not independent of general language skills, but it is still unknown how much language knowledge is necessary to produce a minimally satisfactory story structure at a certain age, or to be able to combine the elements of at least one episode into a GAO-sequence, i.e., to produce a complete episode. Apart from establishing such a threshold, studies of the relationship between macrostructure and microstructure may provide help in identifying children with DLD. In particular, analyses of mental language, such as goals and internal states, could shed light on the specific difficulties of populations with atypical language development. Such analyses would also deepen our understanding of how children develop their reasoning about the inferred parts of events and internal states of story characters.

Last but not least, we want to point out the importance of investigating the cultural dimensions of children's narrative development and supporting the continuous growth of the MAIN network, including the creation of additional language versions as well as fostering further interdisciplinary collaborations by researchers from all around the world.

References

- 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. <https://doi.org/10.1017/S0142716415000466>
- Antonijevic, S., Colleran, S., Kerr, C., & Ní Mhíocháin, T. (2022). Online assessment of narrative macrostructure in adult Irish-English multilinguals. *Frontiers in Psychology*, *13*, Article 916214. <https://doi.org/10.3389/fpsyg.2022.916214>

- Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities*, 55, 301–311. <https://doi.org/10.1016/j.ridd.2016.05.001>
- Blom, E., & Boerma, T. (2020). Bilingual children's lexical and narrative comprehension in Dutch as the majority language. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 198–229). John Benjamins. <https://doi.org/10.1075/sibil.61.07blo>
- Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice: A narrative as diagnostic tool. *International Journal of Language & Communication Disorders*, 51(6), 626–638. <https://doi.org/10.1111/1460-6984.12234>
- 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. <https://doi.org/10.1017/S0142716415000405>
- Bohnacker, U., & Gagarina, N. (Eds.). (2020). *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*. John Benjamins. <https://doi.org/10.1075/sibil.61>
- Bohnacker, U., & Lindgren, J. (2021). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. K. Grohmann (Eds.), *Language Impairment in Multilingual Settings: LITMUS in action across Europe* (pp. 14–46). John Benjamins. <https://doi.org/10.1075/tilar.29.01boh>
- Bohnacker, U., Lindgren, J., & Öztekin, B. (2020). Bilingual Turkish-Swedish children's understanding of MAIN picture sequences: Individual variation, age, language and task effects. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives* (pp. 99–147). John Benjamins. <https://doi.org/10.1075/sibil.61.04boh>
- Bohnacker, U., Lindgren, J., & Öztekin, B. (2022). Storytelling in bilingual Turkish-Swedish children: Effects of language, age and exposure on narrative macrostructure. *Linguistic Approaches to Bilingualism*, 12(4), 413–445. <https://doi.org/10.1075/lab.20057.boh>
- Fiani, R., Henry, G., & Prévost, P. (2020). Narrative comprehension in Lebanese Arabic-French bilingual children. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 32–60). John Benjamins. <https://doi.org/10.1075/sibil.61.02fia>
- Fiani, R., Henry, G., & Prévost, P. (2022). Macrostructure in narratives produced by Lebanese Arabic-French bilingual children: Developmental trends and links with language dominance, exposure to narratives and lexical skills. *Linguistic Approaches to Bilingualism*, 12(4), 446–478. <https://doi.org/10.1075/lab.20035.fia>
- Fichman, S., Walters, J., Armon-Lotem, S., & Altman, C. (2022). The impact of language dominance on Russian-Hebrew bilingual children's narrative production: Microstructure, macrostructure, and Internal State Terms. *Linguistic Approaches to Bilingualism*, 12(4), 509–539. <https://doi.org/10.1075/lab.20036.sve>
- Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: Rasskaz and Erzählung. *Applied Psycholinguistics*, 37(1), 91–122. <https://doi.org/10.1017/S0142716415000430>
- Gagarina, N., & Bohnacker, U. (2022a). A new perspective on referentiality in elicited narratives: Introduction to the Special Issue. *First Language*, 42(2), 171–190. <https://doi.org/10.1177/01427237211062300>
- Gagarina, N., & Bohnacker, U. (2022b). Storytelling in bilingual children: How organization of narratives is (not) affected by linguistic skills and environmental factors. *Linguistic Approaches to Bilingualism*, 12(4), 407–412. <https://doi.org/10.1075/lab.00035.int>

- Gagarina, N., Bohnacker, U., & Lindgren, J. (2019). Macrostructural organization of adults' oral narrative texts. *ZAS Papers in Linguistics*, 62, 190–208. <https://doi.org/10.21248/zaspil.62.2019.449>
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56, 1–140.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, 63, 1–36.
- Gagarina, N., Klop, D., Tsimpli, I. M., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, 37(1), 11–17. <https://doi.org/10.1017/S0142716415000399>
- Gagarina, N., Topaj, N., & Sürmeli, N. (2020). Why do you think the boy would be unhappy if he saw what the cat was eating?: Comprehension of German narratives in Russian- and Turkish-German bilingual children. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 232–268). John Benjamins. <https://doi.org/10.1075/sibil.61.08gag>
- Haddad, R. (2022). *Child bilingualism in Sweden and Lebanon: A study of Arabic-speaking 4-to-7-year-olds*. Acta Universitatis Upsaliensis.
- Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Miękiś, 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. <https://doi.org/10.3389/fpsyg.2017.01444>
- 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. <https://doi.org/10.1017/S0142716415000454>
- Kawar, K., Saiegh-Haddad, E., & Armon-Lotem, S. (2023). Text complexity and variety factors in narrative retelling and narrative comprehension among Arabic-speaking preschool children. *First Language*, 1–25. <https://doi.org/10.1177/01427237221149800>
- Košutar, S., Kramarić, M., & Hržica, G. (2022). The relationship between narrative microstructure and macrostructure: Differences between six- and eight-year-olds. *Psychology of Language and Communication*, 26(1), 126–153. <https://doi.org/10.2478/plc-2022-0007>
- Kunnari, S., & Välimaa, T. (2020). Narrative comprehension in simultaneously bilingual Finnish-Swedish and monolingual Finnish children. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 150–170). John Benjamins. <https://doi.org/10.1075/sibil.61.05kun>
- 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. <https://doi.org/10.1017/S0142716415000442>
- Kuvač Kraljević, J., Hržica, G., & Vdović Gorup, I. (2020). A Comparative Macrostructural Analysis of Narrative Discourse in Children with Typical Language Development and Children with Developmental Language Disorder. *Drustvena Istrazivanja*, 29(3), 453–470. <https://doi.org/10.5559/di.29.3.06>
- Lautenschläger, T., Schneller, K., Kaiser-Kratzmann, J., & Sachse, S. (2021). Erfassung der Erzählfähigkeit mehrsprachig aufwachsender Vorschulkinder: Überprüfung der Testgütekriterien des Verfahrens MAIN (Multilingual Assessment Instrument for Narratives). *Logos*, 29(3), 164–173. <https://doi.org/10.7345/prolog-2103164>

- Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6*. Acta Universitatis Upsaliensis.
- Lindgren, J. (2019). Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language*, 39(4), 412–432. <https://doi.org/10.1177/0142723719844089>
- Lindgren, J. (2022). The development of narrative skills in monolingual Swedish-speaking children aged 4 to 9: A longitudinal study. *Journal of Child Language*, 49(6), 1281–1294. <https://doi.org/10.1017/S030500092100057X>
- Lindgren, J., & Bohnacker, U. (2020). Inferential comprehension, age and language: How German-Swedish bilingual preschoolers understand picture-based stories. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives* (pp. 61–98). John Benjamins. <https://doi.org/10.1075/sibil.61.03lin>
- Lindgren, J., & Bohnacker, U. (2022). How do age, language, narrative task, language proficiency and exposure affect narrative macrostructure in German-Swedish bilingual children aged 4 to 6? *Linguistic Approaches to Bilingualism*, 12(4), 479–508. <https://doi.org/10.1075/lab.20020.lin>
- Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. <https://doi.org/10.1017/S0142716415000429>
- Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2020). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 23(9), 1083–1107. <https://doi.org/10.1080/13670050.2018.1434124>
- Öztekın, B. (2019). *Typical and atypical language development in bilingual Turkish-Swedish preschool children*. Acta Universitatis Upsaliensis.
- Peristeri, E., Andreou, M., Tsimpli, I. M., & Durrleman, S. (2020). Bilingualism effects in the narrative comprehension of children with Developmental Language Disorder and L2-Greek: Links with language, executive function and Theory of Mind. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives* (pp. 298–330). John Benjamins. <https://doi.org/10.1075/sibil.61.10per>
- Pham, G. T., Pruitt-Lord, S., Snow, C. E., Nguyen, Y. H. T., Pham, B., Dao, T. B. T., Tran, N. B. T., Pham, L. T., Hoang, H. T., & Dam, Q. D. (2019). Identifying Developmental Language Disorder in Vietnamese children. *Journal of Speech, Language, and Hearing Research*, 62(5), 1452–1467. https://doi.org/10.1044/2019_JSLHR-L-18-0305
- 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. <https://doi.org/10.1017/S0142716415000417>
- Roch, M., & Hržica, G. (2020). Narrative comprehension by Croatian-Italian bilingual children 5–7 years old: The role of receptive vocabulary and sentence comprehension. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 172–196). John Benjamins. <https://doi.org/10.1075/sibil.61.06roc>
- Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, 21(5), 617–635. <https://doi.org/10.1177/1367006916643528>
- Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative Production in Mandarin-Speaking children: Effects of language ability and elicitation method. *Journal of Speech, Language, and Hearing Research*, 63(3), 774–792. https://doi.org/10.1044/2019_JSLHR-19-00087

*Acquisition of narrative macrostructure:
A comprehensive overview of results from the Multilingual Assessment Instrument for Narratives*

- Tribushinina, E., Irmawati, M., & Mak, P. (2022). Macrostructure in the narratives of Indonesian-Dutch bilinguals: Relation to age and exposure. *Linguistic Approaches to Bilingualism*, 12(4), 540–570. <https://doi.org/10.1075/lab.20015.tri>
- 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. <https://doi.org/10.1017/S0142716415000478>
- Wehmeier, C. M. (2019). Development of narrative macrostructure in monolingual preschoolers in Germany and impact of socio-economic status and home literacy environment. *ZAS Papers in Linguistics*, 62, 52–75. <https://doi.org/10.21248/zaspil.62.2019.443>
- Wehmeier, C. M. (2020). Narrative comprehension and its associations with gender and nonverbal cognitive skills in monolingual and bilingual German preschoolers. In U. Bohnacker & N. Gagarina (Eds.), *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives* (pp. 270–296). John Benjamins. <https://doi.org/10.1075/sibil.61.09weh>
- Yang, W., Chan, A., & Gagarina, N. (2023). Left-behind experience and language proficiency predict narrative abilities in the home language of Kam-speaking minority children in China. *Frontiers in Psychology*, 13, Article 1059895. <https://doi.org/10.3389/fpsyg.2022.1059895>

Appendix

Table 1. Overview of the reviewed MAIN-studies.

Paper	Participants	Ages	Modes/Stories	Language(s)	Prod (SS)	Comp	Analyses
Altman et al. (2016)	Bi English-Hebrew TD (N=19) + DLD (N=12)	5–6	RT BB/BG	English; Hebrew	X*		langs, TD/DLD
Antonijevic et al. (2022)	Bi Irish-English (N=30)	Adults	T Cat/Dog; T BB/BG	English; Irish	X	X	langs, stories
Blom & Boerma (2016)	Mo Dutch TD (N=45) + DLD (N=84)	Long 2yrs T1: 5–6	MS Cat/Dog; T BB/BG	Dutch	X	X	age, TD/DLD
Blom & Boerma (2020)	Mo Dutch (N=45); Bi Turkish-Dutch (N=31); Bi Tarifit-Dutch (N=38)	Long 3yrs T1: 5–6	MS Cat/Dog; T BB/BG	Dutch		X	Mo/Bi, Bi groups, modes, input
Boerma et al. (2016)	Mo Dutch TD (N=33) + DLD (N=33); Bi Dutch TD (N=33) + DLD (N=33)	5–6	MS Cat/Dog; T BB/BG	Dutch	X	X	Mo/Bi, TD/DLD
Bohnacker (2016)	Bi English-Swedish (N=52) ^a	5–7	Telling BB/BG	English; Swedish	X	X	age, langs,
Bohnacker & Lindgren (2021)	Mo Swedish (N=72); ^b Bi English-Swedish (N=52) ^a	Mono: 4–6 Bi: 5–7	T BB/BG; T Cat/Dog (Mono only)	English; Swedish	X		age, langs, Mo/Bi
Bohnacker et al. (2020)	Bi Turkish-Swedish (N=100) ^c	4–7	T Cat/Dog; T BB/BG	Swedish; Turkish		X	age, langs, expressive vocab
Bohnacker et al. (2022)	Bi Turkish-Swedish (N=100) ^c	4–7	T Cat/Dog; T BB/BG	Swedish; Turkish	X		age, langs, expressive vocab, LoE, input
Fiani et al. (2020)	Bi Lebanese Arabic-French (N=48)	4–9	T BB/BG	Lebanese Arabic; French		X	age, langs, dominance, expressive vocab
Fiani et al. (2022)	Bi Lebanese Arabic-French (N=69)	4–9	T BB/BG	Lebanese Arabic; French	X		age, langs, expressive vocab
Fichman et al. (2022)	Bi Russian-Hebrew (N=38)	5–6	RT BB/BG	Hebrew; Russian	X		langs, dominance
Gagarina (2016)	Bi Russian-German (N=57)	3–4 + 6–7 + 8–10	MS Cat/Dog; T BB/BG	German; Russian	X		age, simul/seq
Gagarina et al. (2019)	Mo German (N=30); Mo Russian (N=20); Mo Swedish (N=19)	Adults	T BB/BG	German; Russian Swedish	X		langs
Gagarina et al. (2020)	Bi Russian-German (N=30); Bi Turkish-German (N=27)	long, 3yrs T1: 2–4	MS Cat/Dog; T BB/BG	German		X	age, modes
Haddad (2022)	Bi Arabic-Swedish (N=100)	age 4–7	T Cat/Dog, T BB/BG	Arabic; Swedish	X	X	age, langs, stories

*Acquisition of narrative macrostructure:
A comprehensive overview of results from the Multilingual Assessment Instrument for Narratives*

Haman et al. (2017)	Mo Polish (N=145); Bi Polish-English (N=88)	4–7	T BB/BG; RT Cat/Dog	Polish	X		Mo/Bi, input
Kapalková et al. (2016)	Bi Slovak-English (N=40)	5–6	T BB/BG; RT Cat/Dog	English; Slovak	X	X	langs
Kawar et al. (2023)	Mo Palestinian Arabic (N=30)	5–6	T BB/BG; RT Cat/Dog	Palestinian Arabic; MSA		X	langs, stories
Košutar et al. (2022)	Mo Croatian (N=89)	6 + 8	Telling BB/BG	Croatian	X		age
Kunnari & Välimaa (2020)	Mo Finnish (N=16); Bi Swedish-Finnish (N=16) ^d	5–6	T BB/BG; RT Cat/Dog	Finnish; Swedish		X	langs, Mo/Bi, modes
Kunnari et al. (2016)	Mo Finnish (N=16); Bi Swedish-Finnish (N=16) ^d	5–6	T BB/BG, RT Cat/Dog	Finnish; Swedish	X		langs, Mo/Bi, modes
Kuvač Kraljević et al. (2020)	Mo Croatian TD (N=20) + DLD (N=20)	mean 6;6	T BB/BG, RT Cat/Dog	Croatian	X		TD/DLD, modes
Lindgren (2018)	Mo Swedish (N=72); ^b Bi German-Swedish (N=46); ^c Bi Turkish-Swedish (N=48) ^c	4–6	T Cat/Dog; T BB/BG	Swedish, German, Turkish	X	X	age, langs, Mo/Bi, Bi groups, stories
Lindgren (2022)	Mo Swedish (N=17) ^b	Long 3yrs T1 mean 4;4	T BB/BG	Swedish	X	X	age, stories
Lindgren (2019)	Mo Swedish (N=17) ^b	Long 5yrs T1 mean 4;4	T Cat/Dog	Swedish	X	X	age, stories
Lindgren & Bohnacker (2020)	Bi German-Swedish (N=46) ^c	age 4–6	T Cat/Dog; T BB/BG	German; Swedish		X	age, langs, expressive vocab
Lindgren & Bohnacker (2022)	Bi German-Swedish (N=46) ^c	4–6	T Cat/Dog, T BB/BG	German; Swedish	X		age, langs, expressive vocab, input
Maviş et al. (2016)	Bi Turkish-German (N=49)	2–7	RT/MS Cat/Dog; T BB/BG	Turkish	X	X	age, modes
Otwinowska et al. (2020)	Mo Polish (N=75); Bi Polish-English (N=75)	3–7	RT Cat/Dog; T BB/BG	English; Polish	X	X	langs, Mo/Bi, modes
Öztekin (2019)	Bi Turkish-Swedish (N=102) ^c	4–7	T Cat/Dog, T BB/BG	Swedish; Turkish	X	X	age, langs, stories
Peristeri et al. (2020)	Mo Greek TD (N=30) + DLD (M=30); Bi Albanian-Greek TD (N=30) + DLD (N=30)	6–8	RT Cat/Dog	Greek		X	age, Mo/Bi, TD/DLD, home language history
Pham et al. (2019)	Mo Vietnamese TD/“no risk” (N=45) + “some risk” (N=45) + DLD (N=10)	5	RT Cat	Vietnamese	X		TD/some risk/DLD
Roch et al. (2016)	Bi Italian-English (N=62)	5–7	T BB/BG, RT Cat/Dog	English; Italian	X	X	age, langs, modes

Roch & Hržica (2020)	Bi Croatian-Italian (N=30)	5–7	T BB/BG	Croatian; Italian	X		age, langs, receptive vocab, grammar, AoO, input
Rodina (2017)	Mo Russian (N=16); Mo Norwegian (N=16); Bi Russian-Norwegian (N=16)	4–6	MS Cat/Dog, T BB/BG	Norwegian; Russian	X	X	langs, Mo/Bi
Sheng et al. (2020)	Mo Mandarin Chinese TD (N=21) + “at risk” (N=21)	mean 5;8	T BB/BG; RT Cat/Dog	Mandarin Chinese	X		TD/at risk, modes
Tribushinina et al. (2022)	Bi Indonesian-Dutch (N=32)	5–12	T Cat/Dog (Indonesian); T BB/BG (Dutch)	Dutch; Indonesian	X		age, langs, input
Tsimpli et al. (2016)	Mo Greek TD (N=21) + DLD (N=21); Bi Greek TD (N=15) + DLD (N=21)	mean ≈9	RT Cat/Dog	Greek	X ⁺		Mo/Bi, TD/DLD, composite language ability
Yang et al. (2023)	Bi Kam-Mandarin Chinese (N=55)	5–9	RT Cat/Dog; T BB/BG	Kam	X		age, expressive vocab, grammar
Wehmeier (2019)	Mo German (N=198) ^f	4–5	RT Cat; T BB	German	X	X	age, modes
Wehmeier (2020)	Mo German (N=199); ^f Bi German (N=66)	4–5	RT Cat; T BB	German		X	age, modes

Note. Studies marked with the same superscript letter (e.g. ^a) report results for the same group(s) of participants.¹⁰ Mo = monolingual, Bi = bilingual; TD = typically-developing children; DLD = children with developmental language disorder; T = telling; RT = retelling; MS = model story; BB = Baby Birds, BG = Baby Goats; Prod (SS) = story structure score in narrative production; Comp = story comprehension score; Mo/Bi = comparison of monolinguals and bilinguals; Bi groups = comparisons of bilingual groups; TD/DLD = comparisons of TD children and children with DLD; langs = comparisons of languages; modes = comparisons of elicitation modes; stories = comparisons of stories; vocab = vocabulary; simul/seq = comparison of simultaneous and sequential bilinguals.

* Using a score that only counts goals, attempts, and outcomes (Max = 9)

⁺ Using a score with max =18 (9x2 – two narratives combined, AO=1p, GA/GO=2p, GAO=3p)

¹⁰ This marking is based on information that is either stated explicitly in the cited papers or otherwise known to the authors of the present paper. In some cases, one study reports results for a subgroup of participants that are also included in another study (e.g., the Turkish-Swedish participants in Lindgren, 2018 are a subgroup of the participants in Bohnacker et al., 2020; 2022 and Öztekin 2019).