Article

FIRST LANGUAGE

First Language
1–21

DOI: 10.1177/0142723719844089
journals.sagepub.com/home/fla
SAGE

Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7

Josefin Lindgren
Uppsala University, Sweden; Leibniz-Centre General Linguistics (ZAS), Germany

Abstract

This article reports results from a longitudinal study from age 4 to 7 of comprehension and production of narrative macrostructure in Swedish monolingual children ($N = 17$). *Baby Birds*/*Baby Goats* from the Multilingual Assessment Instrument for Narratives (LITMUS-MAIN) were used to elicit narratives and ask comprehension questions at age 4;4, 5;10 and 7;4. Results showed a steep development from age 4;4 to 5;10 in both comprehension and production of macrostructure, but only some further development in comprehension to age 7;4. For the measures studied, children seem to reach a plateau around age 6. Consistent differences between comprehension and production (higher scores in comprehension) and between stories (higher scores on *Baby Goats*) were found across time points.

Keywords

Longitudinal, macrostructure, Multilingual Assessment Instrument for Narratives (LITMUS-MAIN), narratives, Swedish

Introduction

In recent years, a number of studies have investigated children's language skills using narrative tasks. How children tell stories show their ability to plan and structure longer stretches of complex discourse (Fiestas & Peña, 2004), and to verbalize inferences made about story characters' goals and internal states (emotions) (see Burris & Brown, 2014).

Corresponding author:
Josefin Lindgren, Department of Linguistics & Philology, Uppsala University, Box 365, Uppsala 75126, Sweden.
Email: josefin.lindgren@lingfil.uu.se; lindgren@leibniz-zas.de

Narrative data enable many types of linguistic analysis (e.g. Berman & Slobin, 1994a; Hickmann, 2004; see also the overview in Pavlenko, 2008) and can thus give information both about general processes of language development and of the acquisition of specific phenomena. Narrative tasks can be used to provide an ecologically valid assessment of a child's language proficiency and development (e.g. Botting, 2002).

When telling a story, the speaker needs to structure and present the (story) information in a way that takes the content and the listener's knowledge into account. This requires general language skills, e.g. to use adequate syntactic structures, morphology and vocabulary, and also Theory of Mind, e.g. to understand and verbalize what the story characters think and feel. It has been suggested that narrative competence is the ability to both express the overarching structural organization of the narrative content, the *macrostructure*, and to use specific linguistic structures at the *microstructural* level (e.g. Justice et al., 2006). The macrostructure is thought to be largely language independent. In contrast, the various microstructural aspects, such as the use of a particular (narrative) tense, specific ritualized story openings (e.g. *once upon a time*), or the use of adequate referential, temporal and causal linking devices (so-called cohesive devices; Halliday & Hasan, 1976), are language-specific.

The Multilingual Assessment Instrument for Narratives (LITMUS-MAIN hereafter MAIN; Gagarina et al., 2012, 2015), a picture-based narrative task battery suitable for children aged 4–10, was specifically constructed to assess macrostructure in comprehension and production. Although a relatively new instrument, several studies using MAIN to investigate children's narrative macrostructure have been published, including a special issue of *Applied Psycholinguistics* on this topic in 2016 (e.g. Bohnacker, 2016; Gagarina, 2016; Kunnari, Välimaa, & Laukkanen-Nevala, 2016). However, most of these studies have focused on bilingual children, comparing performance in their two languages. Although several studies investigate age effects, only one (Blom & Boerma, 2016) has investigated development longitudinally. The current article reports a longitudinal study from age 4 to 7 of Swedish monolingual children's comprehension and production of macrostructure in narratives elicited with the MAIN.

This article proceeds as follows. The second section outlines the concept of narrative macrostructure and summarizes findings from earlier studies of preschool children's comprehension and production of narrative macrostructure. The third section states the aim and research questions, and the fourth describes the methodology. The results are presented in the fifth section, followed, finally, by the discussion and conclusion.

Background

Narrative macrostructure: An overview

Narrative macrostructure consists of different components, each with a specific function in the narrative. A number of different narrative models have been proposed to analyse macrostructure (e.g. Labov & Waletzky, 1967; Peterson & McCabe, 1983; Stein & Glenn, 1979). In so-called *story grammar* frameworks (e.g. Mandler & Johnson, 1977; Stein & Glenn, 1979; Stein & Policastro, 1984), narratives consist of a *setting* (time and place of the events, often also introductions of protagonists) and an *episodic system*.

Table 1. Macrostructural components in the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015).

Component	Description
Setting	Time and place of the events
Internal State as Initiating Event (IS 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 (IS as R)	What does the character feel (in response to the outcome)?

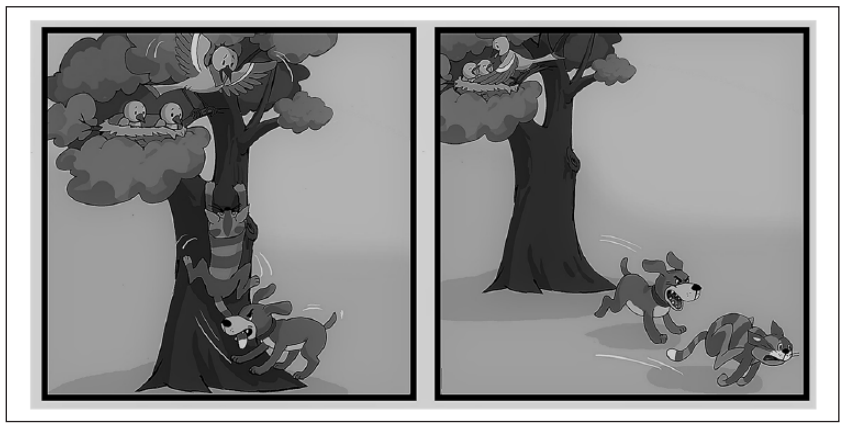

Figure 1. Small-scale, black-and-white version of pictures 5–6, *Baby Birds*, Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015).

Episodes are seen as *goal-based*, viewing the goal as 'the most critical piece of information, for story knowledge is basically organized around the goal of the protagonist' (Stein & Policastro, 1984, p. 118). Which components are part of the episode varies somewhat between models. Adapted versions of the Stein and Glenn (1979) story grammar model have been widely used in studies of child language (e.g. Fiestas & Peña, 2004; Trabasso & Nickels, 1992).

The model underlying MAIN (Gagarina et al., 2012, 2015), the narrative instrument used in the current study, is also based on (goal-based) story grammar models, and on Westby's (2012) decision-tree model of narrative structure (see below). The pictures used as stimulus materials in MAIN were specifically constructed to elicit utterances containing the components shown in Table 1.

To give an example of how the components are included (i.e. depicted) in the MAIN pictures, Figure 1 shows pictures 5–6 from the story *Baby Birds.* In these two pictures, attempt ('The dog bites the cat's tail and pulls it down the tree'), outcome ('The dog chases the cat away') and internal state as reaction ('The cat is scared' or 'The birds are happy')

Table 2. Episodic structures (following Westby, 2012).

Type of sequence	Macrostructural components
Action/reaction sequence	Attempt + Outcome (AO)
Incomplete episode	a. Goal + Attempt (GA) b. Goal + Outcome (GO)
Complete episode	Goal + Attempt + Outcome (GAO)

for the story's third episode are shown. Narrative macrostructure and story content more generally can thus not only be told verbally, but also be presented visually, e.g. shown in pictures (as in the present study) or in video-clips (cf. Bishop, 1997, chap. 7; Bishop & Adams, 1992; Lynch et al., 2008; Paris & Paris, 2003; van den Broek et al., 2005).

In addition to analysing the production of different macrostructural components, the narrative's structural complexity can be assessed. Based on Westby's (2012, p. 211) decision-tree model, narrative episodes can be classified according to their structural complexity as action/reaction sequences (attempt–outcome sequences), which contain only an attempt and its outcome, incomplete or abbreviated episodes, which contain a goal but lack either the outcome (goal–attempt sequences) or the attempt (goal–outcome sequences), and complete episodes (goal–attempt–outcome sequences) (see Gagarina et al., 2012, pp. 11–12), as shown in Table 2.

Complete episodes indicate a high level of structural complexity, i.e. a more well-formed narrative (cf. Gagarina et al., 2012, p. 11, 2015; Stein & Glenn, 1979; Westby, 2012). To tell a picture-based narrative using only action/reaction sequences means including only actions that are visible in the pictures. For incomplete and complete episodes, it is also necessary to verbalize inferences about the characters' goals drawn from the pictures. Attributing goals to characters and overtly expressing these is more complex than telling events shown in pictures. Telling a narrative with a specific level of macrostructural complexity, i.e. using specific sequence types, thus shows a certain level of narrative competence, representing different developmental stages (cf. Gagarina et al., 2015; Westby, 2012).

Children's production of macrostructure

What is included in measures of macrostructure as well as the method used for eliciting narratives from children (e.g. telling vs retelling, with vs without support from pictures) varies between studies, making it somewhat difficult to generalize the results. However, findings across studies and languages show that narrative macrostructure develops extensively from age 3 to 7 (e.g. Berman & Slobin, 1994b; Bohnacker, 2016; Lindgren, 2018; Pearson, 2002; Schneider, Hayward, & Dubé, 2006; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992). Beginning at age 3–4, children's narratives contain not only *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), but also *action sequences*, loosely linked with simple connectives such as *and* or *then* (e.g. Berman &

Slobin, 1994b; Peterson & McCabe, 1983; Westby, 2012). Children also start producing *reaction sequences* (Stein & Albro, 1997), in which there is causal linking but no mention of characters' goals. Around age 5, goals appear in the children's narratives, and the proportions of *complete episodes*, i.e. goal–attempt–outcome sequences, increase as children grow older (e.g. Trabasso & Nickels, 1992). Trabasso et al. (1992) found that not until age 9 did proportions of goals reach the same level as in adults. In contrast, attempts and outcomes were relatively frequent already at age 4 (Trabasso et al., 1992). Older children thus generally produce more complex episodic structures. Most studies were cross-sectional, and thus did not investigate narrative development in the same children.

Generally, little is known about narrative development in Swedish-speaking children. Apart from the studies using MAIN described below, a number of studies of Swedish-speaking children's narratives have been published (Nordqvist, 2001; Reuterskiöld, Hansson, & Sahlén, 2011; Reuterskiöld Wagner, Sahlén, & Nettelbladt, 1999; Strömqvist & Day, 1993; Viberg, 2001). However, only Nordqvist (2001) analysed narrative macrostructure. Following the seminal study by Berman and Slobin (1994b) on the Frog story (Mayer, 1969) narratives from monolingual children aged 3–4, 5, 7 and 9 and adults speaking English, German, Hebrew, Spanish and Turkish, Nordqvist (2001) analysed production of three core plot components (onset, unfolding, resolution) in the same type of narratives told by 3-, 4-, 5-, 9- and 15-year-old monolingual Swedish children and adults ($N = 84$). She found development with age, especially between age 4 and 5, for each of the three components (Nordqvist, 2001, pp. 201–203).

The results from previous studies of narratives elicited with MAIN are similar to earlier studies using other stimuli. A range of languages have been investigated using MAIN, but focus has mainly been on the performance of bilingual children either in their two languages or, in a few cases, compared to monolinguals, e.g. Swedish monolinguals, Swedish–German and Swedish–Turkish bilinguals (Lindgren, 2018), Swedish–English bilinguals (Bohnacker, 2016), Finnish monolinguals and Finnish–Swedish bilinguals (Kunnari et al., 2016), Dutch monolinguals (Blom & Boerma, 2016), Dutch monolinguals and bilinguals (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016), Polish monolinguals and Polish–English bilinguals (Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, & Haman, 2018), Russian–German bilinguals (Gagarina, 2016), Turkish–German bilinguals (Maviş, Tunçer, & Gagarina, 2016) and early L2 learners of English, with L1 Italian (Roch, Florit, & Levorato, 2016).[1]

Previous studies using MAIN analysed the production of macrostructural components using the same composite story structure (or macrostructure production) score (see subsection on scoring below). For example, Lindgren (2018) analysed narratives by Swedish monolinguals ($N = 72$), and in both languages of Swedish–German bilinguals ($N = 46$) and Swedish–Turkish bilinguals ($N = 48$) aged 4–6. She found a clear effect of age on the inclusion of macrostructural components in all three language groups in Swedish; 6-year-olds had higher story structure scores than 5-year-olds, who in turn performed better than 4-year-olds. Bohnacker (2016) found that Swedish–English bilingual 6- to 7-year-olds ($N = 33$) performed better than 5-year-olds ($N = 19$), irrespective of language. Kunnari et al. (2016) investigated 16 Finnish monolinguals and 16 Finnish–Swedish bilinguals (growing up in Finland) aged 5;0–6;7 and found a significant

improvement with age (in months) in the story structure score. Gagarina (2016) studied both languages of 58 Russian–German bilinguals growing up in Germany, analysing narratives from 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 the story structure score in both languages, but there were no differences between the older groups.

Of the studies using MAIN, only Blom and Boerma (2016) investigated development longitudinally. In a large-scale study of Dutch monolinguals with typical language development (TD, $N = 45$) and with language impairment (LI, $N = 84$), they compared the children's story structure scores at age 5–6 ($M = 5;9$) and age 6–7 ($M = 6;9$). Interestingly, their results showed a development with age in the LI group, but not in the TD group.

A number of the studies have also investigated macrostructural complexity. The specific measures used to analyse complexity vary between studies. Some studies (Gagarina, 2016; Roch et al., 2016) assessed the child's best performance, i.e. they assigned a complexity score to the child based on the most complex sequence used in the narrative. Other studies classified each episode according to its type of sequence and 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 all episodes (Maviş et al., 2016), compared the proportions of different types of sequences used by different age- and language groups (Lindgren, 2018), or investigated the frequency of GAO sequences (Bohnacker, 2016).

Gagarina (2016) found that Russian–German bilingual preschoolers (aged 2;7–4;4) produced few GAOs in either language. In German, she only found a difference between the preschoolers and the two older groups (grade 1, grade 3) in the proportion of children producing at least one GAO. In Russian, a higher proportion of the grade 3 children than the grade 1 children produced one or more GAO. Kunnari et al. (2016) found a clear development with age (in months) in proportions of episodes containing a goal in both Finnish and Swedish. Bohnacker (2016) found that Swedish–English 5-year-olds produced only 7% GAOs, and only slightly more at age 6–7 (12%). She also found that 21% of the narratives of the younger group contained at least one GAO; the corresponding figure for the 6–7-year-olds was 35%. In analyses of Swedish data from two narrative tasks (MAIN *Cat*/*Dog* and *Baby Birds*/*Baby Goats*) from mono- and bilingual Swedish-speaking children ($N = 166$), Lindgren (2018) found significant effects of age both on production of macrostructural sequences (vs not producing any type of sequence) and on GAOs; 6-year-olds produced more sequences and more GAOs than 5-year-olds, who in turn performed better than 4-year-olds. In the majority of the episodes (60%) in *Baby Birds*/*Baby Goats*, the monolingual 6-year-olds produced AO sequences; GAOs (11%) were comparably rare (GA/GO sequences were the least frequent type). The 4-year-olds produced only 6% GAOs. Out of all 72 monolingual children aged 4–6, 20% produced at least one GAO sequence in their narrative.

Differences between the two stories, *Baby Birds* and *Baby Goats*, have only been investigated in two studies (Bohnacker, 2016; Lindgren, 2018). There was no significant difference between the two stories in story structure score for the 166 Swedish-speaking children in Lindgren (2018) or for the 52 Swedish–English bilinguals in Bohnacker (2016), indicating that the two stories may be equivalent in terms of macrostructure

production. However, macrostructural complexity has not been investigated for the stories separately.

The results from studies using MAIN to assess macrostructure show that production of macrostructure develops in the preschool period, but that performance may not consistently increase in the early school age. GAOs were relatively rare in MAIN narratives from children aged 4–7.

Children's comprehension of macrostructure

When comprehending a story (or any other type of discourse), we create a mental representation of the story content (e.g. Stein & Glenn, 1979), i.e. of the narrative macrostructure, based on information provided in the story and our background knowledge. As mentioned above, the story information may be presented orally or visually. Comprehending a story is not the same as recalling story events, but rather understanding the relations between events, including the reasons for a character's actions, i.e. understanding the plot of the story (van den Broek et al., 2005, pp. 118, 126). This requires us to draw inferences (Hayward, Schneider, & Gillam, 2009; Stein & Glenn, 1979).

The ability to understand an underlying story schema (see above), including goal planning, is often seen as a prerequisite for being able to produce full episodic structures in narratives (cf. Burris & Brown, 2014; Stein & Glenn, 1979; Trabasso & Rodkin, 1994). It is therefore important to investigate story comprehension in addition to production. However, far fewer studies have investigated narrative comprehension than production. In fact, comprehension is not always clearly separated from production; sometimes story production methods, such as narrative retell, are listed as probing narrative comprehension (cf. Burris & Brown, 2014). When narrative comprehension is measured directly, it is mostly in the form of probe questions, often such that require the child to make inferences (e.g. Bohnacker, 2016; Lindgren, 2018; Stein & Glenn, 1979; Trabasso et al., 1992), asked either after a child has heard and retold a story or after a child has told a story from pictures (as in the current study).

Earlier studies point to children's comprehension of narrative structure developing earlier than their ability to verbalize it in their narratives. In a classic study, Stein and Glenn (1979) analysed answers to narrative comprehension questions (most of the questions targeted inferences) by 24 monolingual English 6- and 10-year-olds. They found that 6-year-olds had overall good narrative comprehension, even of goals and internal states, story components which the children rarely expressed in their narratives. Similarly, Trabasso et al. (1992) found that English monolingual 4-year-olds, who did not spontaneously produce goals of story characters in their narratives, could give reasons why a character performed an action when explicitly probed. In a recent longitudinal study using the Test of Narrative Language (TNL; Gillam & Pearson, 2004), Gibson, Peña, and Bedore (2018) analysed English narratives from Spanish–English bilinguals with typical language development ($N = 20$) and with language impairment ($N = 20$) at age 6 and age 7. In the TNL story generation task with multiple pictures (the task most similar to the one used in the current study), the typically developing children performed consistently better in comprehension than in production across time points.

Out of the studies using MAIN cited above, only five (Boerma et al., 2016; Bohnacker, 2016; Lindgren, 2018; Maviş et al., 2016; Roch et al., 2016) report results for the MAIN *Baby Birds*/*Baby Goats* comprehension questions (used in the telling mode).[2] Out of these, only the studies by Bohnacker (2016) and Lindgren (2018) compared the children's performance in comprehension and production. Bohnacker (2016) found that Swedish–English bilinguals performed better in comprehension than in production, and although there was a development from age 5 to age 6–7 in comprehension, comprehension was already at a very high level at age 5. Even though not even the older group produced many goals in their narratives, all children understood them well when probed. Lindgren (2018) found a significant increase with age in the comprehension scores of Swedish monolinguals, Swedish–German bilinguals and Swedish-Turkish bilinguals. While comprehension was good already at age 4, production was still relatively poor at age 6; across age groups these Swedish-speaking children performed significantly better in comprehension than in production. However, the measures used for comprehension and production in these two studies were not completely equivalent, i.e. they did not include identical macrostructural components. The current study contributes to our understanding of the relationship between narrative comprehension and production by comparing children's performance on the same measure for comprehension and production.

Concerning differences between stories, Lindgren (2018) found for 166 Swedish-speaking children that comprehension scores were significantly higher on *Baby Goats* than on *Baby Birds*, whereas Bohnacker (2016) found no differences between the stories for 52 Swedish–English bilinguals. However, in a reanalysis of the comprehension data from Bohnacker (2016) and the monolingual data from Lindgren (2018), Bohnacker and Lindgren (in press) did find differences between the stories also for these Swedish–English bilinguals, but only in English. There are thus indications that *Baby Birds* and *Baby Goats* may not be completely equivalent in comprehension.

Earlier studies thus show age effects on children's ability to comprehend narrative macrostructure and a consistent gap between comprehension and production across age groups. The current study investigates this longitudinally, for the first time in a study using MAIN.

Aim and research questions

The aim of the current study is to investigate the development of the comprehension and production of narrative macrostructure from age 4 to age 7 in Swedish monolingual children. Additionally, the study analyses effects of the specific story told by the children. The following three research questions were asked.

1. How do the children's comprehension and production of narrative macrostructure develop between age 4 and 7?
2. How does the performance in narrative comprehension compare to production?
3. Is there a difference in performance between the two stories used as stimulus materials (*Baby Birds* vs *Baby Goats*)? If there is a difference, is it the same at all three points of testing?

Table 3. Participants ($N = 17$).

	T1	T2	T3
Mean age (SD)	4;4 (0;3)	5;10 (0;3)	7;4 (0;3)
Age range	4;0–4;8	5;5–6;2	6;11–7;8

Note. T1 = first testing, T2 = second testing, T3 = third testing.

Methods

Participants

The participants were 17 Swedish monolingual children (10 girls, 7 boys).[3] Data were collected from the children at three time points with 18 months in between, with the first at age 4 and the third at age 7. The first testing (T1) took place in September/October 2014, the second (T2) in March/April 2016, and the third (T3) in September 2017. T1 and T2 took place in the children's preschools and T3 in their schools, where they had just started grade 1. The participants were recruited by staff from two preschools at T1; at T3, they attended 13 different schools. Parents signed consent forms and filled in brief background questionnaires at each of the three time points. All children lived in Gävle or Uppsala, two larger Swedish cities, came from mid- to high-SES backgrounds as measured by parental education,[4] and had typical language development according to parental report. Table 3 gives an overview of the participants with respect to their ages at the three testing points.

Materials

The children told narratives and answered comprehension questions to *Baby Birds* or *Baby Goats* from the MAIN (Gagarina et al., 2012, 2015). These tasks, suitable for children aged 4–10, consist of picture sequences with six pictures depicting a three-episode story and 10 standardized comprehension questions, targeting characters' goals and internal states. In order to answer the comprehension questions, the child needs to draw inferences from the events shown in the pictures. Each episode shown in the pictures contains a GAO sequence for a specific character (cf. Stein & Glenn, 1979). *Baby Birds* and *Baby Goats* have been carefully constructed to be comparable (see Gagarina et al., 2012); the stories are parallel in terms of length and story grammar components. Both contain five story characters. In addition to the picture sequences, the 10 comprehension questions (including scoring guidelines for the children's answers) and a standardized procedure for administering the task (see below), MAIN also contains a protocol for assessing production of macrostructural elements (story structure components) and narrative (macrostructural) complexity (see below).

Procedure

The author administered the narrative task to all children. The children told the same story, following an identical procedure, at all three time points, as part of a larger battery

of tasks (see Lindgren, 2018, chap. 3 for details). The *Baby Birds*/*Baby Goats* task was administered in the middle of the session, after the children had told another story (MAIN *Cat*/*Dog*) and answered its comprehension questions.[5] Nine children told *Baby Birds* and eight children told *Baby Goats*. There was no significant difference in age at T1 between the children who told *Baby Birds* (M = 4;6) and those who told *Baby Goats* (M = 4;3), $t(15) = 1.951$, $p = .07$.[6] The standardized procedure for administering MAIN in the telling mode including comprehension questions was used (see Gagarina et al., 2012). This procedure is as follows. The child chooses one envelope from three placed on the table and takes out the pictures inside. All envelopes contain the same story, but the child is led to believe that the experimenter does not know which story the child will tell. The child looks at the six pictures, presented as a fold-out strip. The experimenter folds back the pictures so that only the first two are visible to the child and asks the child to begin the story. When the child has finished telling about the first two pictures, the next two pictures are unfolded and finally the last two, so that all six pictures are visible to the child. Throughout the storytelling, the experimenter gives only minimal prompts (e.g. *aa, mm, and then?*). At no point before or during the storytelling is the experimenter allowed to look at the pictures. When the child has finished telling the story, the pictures are placed on the table so that they are visible to both child and experimenter while the comprehension questions are asked.

Scoring of macrostructure in comprehension and production

All narratives and answers to comprehension questions were audio and video recorded and transcribed in CHAT format (MacWhinney, 2000). There were no missing data; all children told a story in all three testings, and all comprehension questions were asked and answered. The data thus consisted of 51 narratives and 510 comprehension questions.

For each of the 10 comprehension questions, the correct answer is awarded 1 point (maximum score = 10 points). Three questions, one for each episode, probe the child's understanding of the goals of the main story characters (e.g. *Why does the mother bird fly away?* for Episode 1 goal in *Baby Birds*). Three further questions target character's internal states (e.g. *How does the cat feel?* (in pictures 5–6) for Episode 3 internal state as reaction in *Baby Birds*), with three follow-up why-questions probing the rationale behind the character's internal states (*Why does the cat feel like that?*). The final question requires understanding of the overall plotline (*Who does the mother bird like best, the cat or the dog? Why?*). The comprehension questions thus assess the child's ability to verbalize inferences they have drawn from the story content shown in the pictures.

The MAIN scoring protocol for the production of macrostructure awards points for the successful production of setting (2 points, one for time and one for place), 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 (3 × 5 points), yielding a maximum macrostructure production score of 17 points per narrative. Table 4 gives an overview of the macrostructural components in the setting and Episode 1 of *Baby Goats*.

Table 4. Overview of macrostructural components in MAIN with constructed examples, *Baby Goats*, setting and Episode 1.

Component	*Baby Goats*
Setting	Once upon a time… (*time*) …by a lake (*place*)
Episode 1	
IS as IE	The mother goat saw that the baby goat was in danger
Goal	She wanted to rescue it
Attempt	She ran down into the water…
Outcome	…and pushed the baby out of the water
IS as R	The mother goat was happy

Note. IS = internal state, IE = initiating event, R = reaction.

Each child received one comprehension and one production score per narrative task (i.e. per time point). Additionally, the narratives were scored for macrostructural complexity, focusing on the production of sequences of core macrostructural components (goal, attempt, outcome) within an episode. The child's production for each episode was classified as either no sequence (only one or none of the components goal, attempt and outcome were produced within the episode),[7] attempt–outcome sequence (AO, action/reaction sequence), goal–attempt/goal–outcome sequence (GA/GO, incomplete episode), or complete episode, i.e. goal–attempt–outcome sequence (GAO).

As explained above, the MAIN comprehension and production scores do not include the same narrative components. To make a valid comparison between comprehension and production it was therefore necessary to create modified comprehension and production scores that only include the five components targeted in both comprehension and production (Episode 1 internal state as initiating event, Episode 1 goal, Episode 2 goal, Episode 3 goal, Episode 3 internal state as reaction). The maximum for the modified comprehension and production scores is 5 points. While the standard macrostructure production score includes also settings, attempts and outcomes, the modified production score only includes goals and internal states, i.e. those components thought to be more difficult to produce.

The author scored all narratives and all answers to the comprehension questions, following the scoring decisions of the MAIN manual (Gagarina et al., 2012). However, there was substantial variation between individual children in how the narrative content and answers to the comprehension questions were phrased, including a number of cases for which the manual did not provide a scoring decision. For this reason, detailed scoring guidelines including scoring decisions and the rationale behind them for a more diverse range of cases as well as general scoring principles were developed within the BiLI-TAS project (Bohnacker, 2013), with which the present study was affiliated. These guidelines were based on extensive discussions of unclear cases found in the answers and narratives of 286 mono- and bilingual children in the project group (including the author). The author checked the scoring thoroughly for consistency against these guidelines.

Table 5. Comprehension scores by time point (Max = 10 points).

	T1 (4;4)	T2 (5;10)	T3 (7;4)
Mean	5.3	8.1	8.9
SD	2.5	1.6	1.0
Range	1–9	5–10	7–10

Two principles were central for scoring narrative production. First, focus was on the action or emotion described by the child and whether this was realized verbally in a clear enough manner. Non-verbal aspects such as pointing or iconic gestures were not considered. The definition of clear enough was that an action/emotion was understandable and could be linked to a specific character and point in the story, meaning that the use of relatively specific verbs as well as the inclusion of appropriate agents/patients were crucial to receive a point. Second, a principle of lenient scoring of reference was used to avoid that the ability to correctly introduce and refer back to story characters and objects would determine a child's production score. It was not necessary to employ adultlike use of reference to score on the macrostructural components, as long as the utterance fulfilled the requirements stated in the first principle.

Results

Comprehension of macrostructure

Table 5 shows the children's scores on the 10 comprehension questions.

At age 4;4, the children scored around 50% correct on the comprehension questions. One and a half years later (age 5;10), their accuracy was at 80%. At age 7;4, they reached almost 90% correct. The children's ability to comprehend characters' goals and emotions, as well as the general plotline, thus appears to develop steeply between age 4 and 6, but after that reaches a plateau, relatively close to the maximum score. A repeated-measures ANOVA with Time (of testing) as within-child factor, and Story as between-child factor, supports this interpretation: the main effect of Time was highly significant, $F(2, 30) = 27.094$, $p < .001$, $\eta^2_p = .644$, but subsequent post-hoc tests (Bonferroni) revealed that there was a highly significant difference between ages 4;4 and 5;10 ($p < .001$), but that the increase in scores from age 5;10 to 7;4 was not statistically significant ($p = .235$). Additionally, performance on the two stories, *Baby Birds* and *Baby Goats*, differed significantly from each other, $F(1, 15) = 7.444$, $p = .016$, $\eta^2_p = .332$, with higher scores on *Baby Goats* than on *Baby Birds*. Figure 2 shows the mean scores for the three time points by story. Although the difference between the stories seems to be somewhat larger in the first testing, the interaction between Time and Story was not significant, $F(2, 30) = 1.725$, $p = .195$, $\eta^2_p = .103$; the gap between the performance on the two stories was the same at the three time points. Across stories, variation between individual children was very large at T1, smaller at T2, and very small at T3.

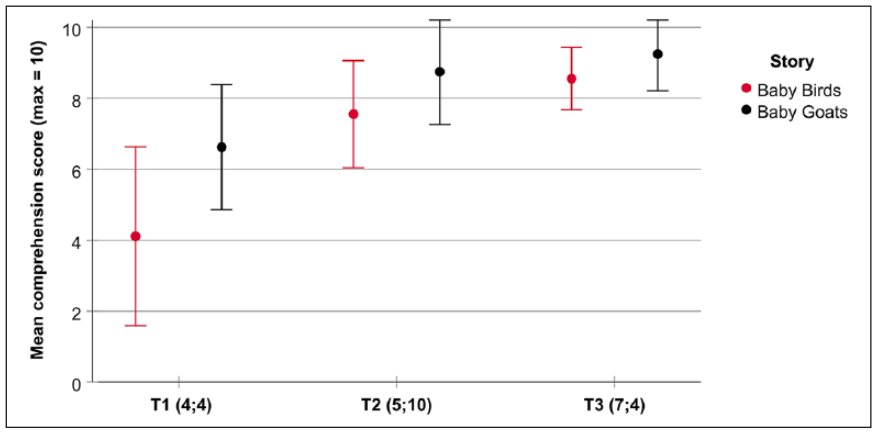

Figure 2. Mean comprehension scores by time point and story. Max = 10 points. Error bars show ±1 SD.

Table 6. Macrostructure production scores by time point (Max = 17 points).

	T1 (4;4)	T2 (5;10)	T3 (7;4)
Mean	5.2	7.4	8.0
SD	2.3	2.0	1.7
Range	2–10	4–11	5–12

Production of macrostructure

Macrostructure production score. In Table 6, the macrostructure production scores (story structure score) for the three time points are shown.

At age 4, the children produced around 30% of the scored macrostructural components, a number which had risen to close to 50% at age 7. The performance in production was thus relatively low even at age 7. Just as was the case in comprehension, the scores of the two later time points were close to each other. The repeated-measures ANOVA for production did indeed show the same result as for comprehension, namely a significant effect of Time, $F(2, 30) = 9.271$, $p = .001$, $\eta^2_p = .382$, with a significant development from age 4;4 to 5;10 ($p < .035$), but no difference between age 5;10 and 7;4 ($p = .911$), and a significant effect of Story, $F(1, 15) = 9.212$, $p = .008$, $\eta^2_p = .380$, with higher scores on *Baby Goats* than on *Baby Birds*. The interaction between Time and Story was not significant, $F(2, 30) = .062$, $p = .940$, $\eta^2_p = .004$; the difference between the two stories was constant across time points, as shown in Figure 3.

Macrostructural complexity. In Figure 4, results for macrostructural complexity, i.e. the production of different types of sequences of the core macrostructural components (goals, attempts, outcomes) are shown.

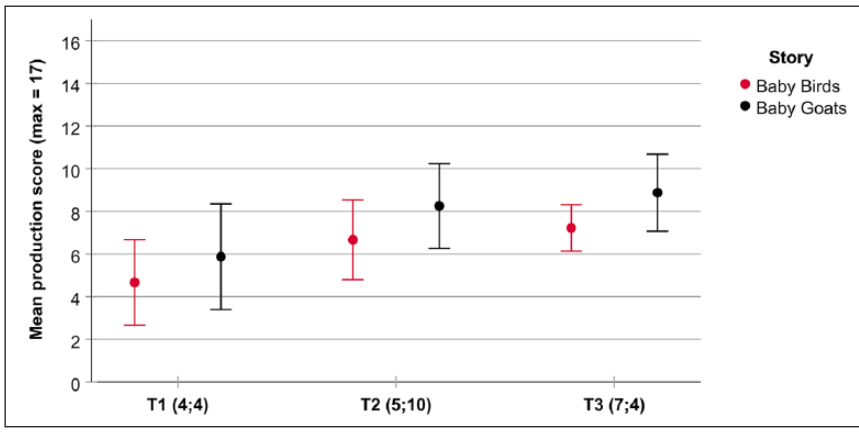

Figure 3. Mean macrostructure production scores by time point and story. Max = 17 points. Error bars show ±1 SD.

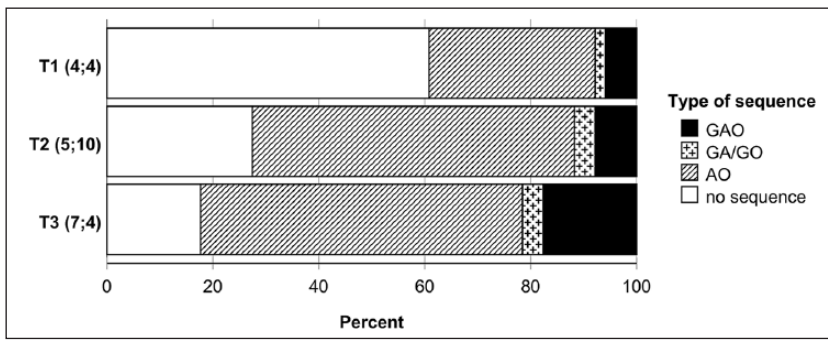

Figure 4. Types of different sequences of core macrostructural components out of all episodes (N = 153), by time point.

As seen in Figure 4, there was a significant overall association between the proportions of different types of sequences used and time of testing, $\chi^2(6, N = 153) = 24.822$, $p < .001$. Post-hoc tests (Bonferroni-corrected) showed that the children produced a significantly higher proportion of episodes that did not contain any sequence of the three core components ('no sequence') at age 4;4 (60.8%) compared with ages 5;10 (27.5%) and 7;4 (17.6%); the difference between the two later time points was not significant. The children also produced significantly fewer AO sequences at age 4;4 (31.4%) compared with age 5;10 (60.8%) and age 7;4 (60.8%). The production of GA/GO sequences was low and almost identical at all three time points (see Figure 4). Although at age 7;4, the children produced 17.6% GAOs, compared with 7.8% at age 5;10 and 5.9% at age 4;4, this difference did not reach significance. There was also no significant difference between the three testings in the proportion of children who produced at least one GAO,

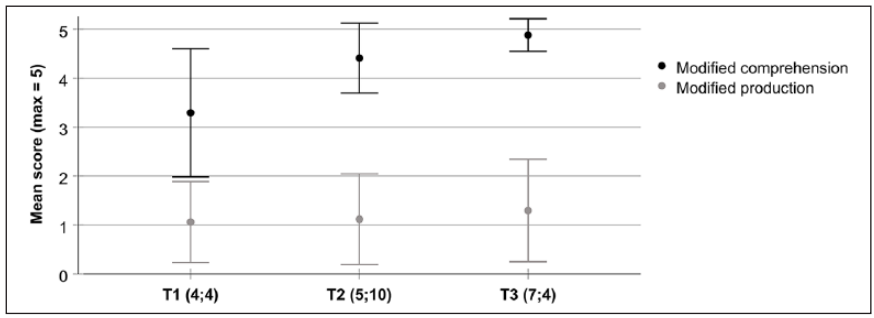

Figure 5. Modified comprehension and production scores by time point. Max = 5 points. Error bars show ±1 SD.

$\chi^2(2, N = 51) = 4.923, p = .085$, despite the fact that 47.1% (8/17) of the children did so at T3, compared with only 17.6% (3/17) at T1 and T2. This could be an effect of the relatively small number of children.[8] Interestingly, there was no difference between *Baby Birds* and *Baby Goats* in macrostructural complexity, $\chi^2(3, N = 153) = 3.519, p = .318$; although the children scored fewer points on macrostructure in *Baby Birds* than in *Baby Goats*, narratives told to the two stories thus had equally levels of complexity.

Comparison of comprehension and production

Finally, comprehension and production were compared using the modified scores which includes three goals and two internal states. The results are shown in Figure 5. The repeated-measures ANOVA with Time (of testing) and Modality (comprehension vs production) as within-child factors showed significant mains effects of both Time, $F(2, 32) = 7.714, p = .002, \eta^2_p = .325$, and Modality, $F(1, 16) = 354.613, p < .001, \eta^2_p = .957$, but most importantly, a significant interaction between Time and Modality, $F(2, 32) = 5.570, p = .008, \eta^2_p = .258$. Subsequent simple effects analyses revealed that comprehension scores were higher than production scores at all three time points (*p*s < .001). However, while there was a significant effect of Time in comprehension, $F(2, 15) = 14.239, p < .001$, with post-hoc tests showing significant development both from age 4;4 to age 5;10 ($p = .006$), and from age 5;10 to 7;4 ($p = .027$), this was not the case in production, $F(2, 15) = .214, p = .810$. The difference between the modified comprehension and production scores was thus larger at age 7 than at age 4. Additionally, variation in narrative comprehension was much smaller at the later age; in fact, at T3, all children scored close to ceiling on modified comprehension. In production, variation was substantial at all time points.

Discussion and conclusion

This article presented results from a longitudinal study on comprehension and production of narrative macrostructure in Swedish monolinguals ($N = 17$) aged 4 to 7. Narratives

were elicited and comprehension questions asked at three time points (mean ages 4;4, 5;10 and 7;4) using the picture-based *Baby Birds*/*Baby Goats* task from the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2012, 2015). The comprehension questions targeted the child's ability to verbalize inferences about character's goals and internal states drawn from the story shown in the pictures. The three research questions concerned the development with age for comprehension and production (1), how performance in comprehension compared to production (2) and whether there was any difference between the two stories (3).

Regarding the first question, the results showed age development in both comprehension and production, in line with earlier studies (e.g. Bohnacker, 2016; Gagarina, 2016; Kunnari et al., 2016; Lindgren, 2018). However, while there was a significant development from age 4;4 to age 5;10, with a large improvement in scores, the scores of ages 5;10 and 7;4 did not differ significantly. This is similar to the results from Blom and Boerma (2016), the only other longitudinal study using MAIN, where there was no improvement in typically developing Dutch monolingual children's scores from age 5;9 to 6;9, but differs from the results of the cross-sectional study of Swedish–English bilinguals by Bohnacker (2016), who found significant differences between 5-year-olds and 6- to 7-year-olds in comprehension and production. The question remains if the reason for the apparent lack of development from age 5;10 to 7;4 is that the children's narrative macrostructure does not improve or if the measures are not sensitive enough to detect further development. In fact, this seemed to be the case in comprehension – the modified comprehension score, including only goals and internal states, did show a significant difference between the two later testing points. It could be the case that the children also improve in production, but in a way that is not assessed by the macrostructure production score.

Narratives produced at the two later time points did not only score higher on a total score of macrostructure but also showed a higher level of macrostructural complexity. At age 4;4, the narratives generally had a low level of complexity, with the children only producing sequences of the core macrostructural components (goals, attempts, outcomes) in around 40% of the episodes; in the other 60%, they only produced one of the three components. At the later time points, the children's narratives contained higher proportions of sequences, but again there was no significant difference between age 5;10 and 7;4. Macrostructural complexity, as operationalized in the current study, did thus not show any development between ages 6 and 7. A significant increase between age 4;4 and the two later time points was found for attempt–outcome sequences, but not for sequences containing goals (GA/GO, or GAO). Interestingly, across time points, the children more commonly produced full episodes, GAO sequences, than abbreviated episodes, GA/GO sequences. This indicates that when children do produce a goal, which still is relatively rare even at age 7, they mainly do so *in addition to* both attempt and outcome. Children thus seem to first begin to include all attempts and outcomes in their narratives and then later sometimes add the goals. The proportions of GAOs are comparable with earlier studies of similar age groups (e.g. Bohnacker, 2016; Lindgren, 2018); the relatively low proportions indicate that children have not yet mastered narrative macrostructure in production at this age.

At age 6–7, these Swedish-speaking children showed signs of mastery of narrative comprehension, but still had relatively low scores in production, in line with earlier studies (Bohnacker, 2016; Lindgren, 2018). When comprehension and production were compared, using the exact same measures in both modalities (modified comprehension and production scores containing only goals and internal states), the difference between them was highly significant (see Figure 5). The children's understanding of goals and internal states remained at a significantly higher level than their production throughout the studied ages. As mentioned above, the modified measure was sensitive to development in the children's comprehension from age 5;10 to 7;4, which the original comprehension score was not. This later development became clear when only goals and (certain) internal states were considered. On the other hand, the modified production measure showed no development with age, and thus hides the difference between narratives told at age 4;4 and at the later time points, a difference that is clear when the narratives are analysed using the original MAIN story structure score (see Table 6). What is included in a measure influences whether development can be seen.

To turn to the final research question, whether there was any difference between the two stories, there was a significant effect of story. In both comprehension and production and across time points, scores for *Baby Goats* were higher than for *Baby Birds*. However, narratives produced with the two stories had an equal level of macrostructural complexity. Earlier cross-sectional studies using MAIN *Baby Birds*/*Baby Goats* found differences in comprehension, but not in production (Bohnacker & Lindgren, in press; Lindgren, 2018). The reason for the difference between studies is unclear. In the current study, there were no apparent differences between the children who told the *Baby Birds* story and those who told *Baby Goats*: they were equally old, came from the same type of background, and had comparable scores on a vocabulary task. The stories have comparable episodic structures, number and types of characters, and comprehension questions. Future studies should investigate these stimulus materials in detail to determine the reason for the differences in performance.

What do the results then tell us about children's receptive and expressive narrative abilities? First, children aged 4–7 primarily tell stories that include the macrostructural components that are clearly shown in the pictures such as attempts and outcomes (cf. Bohnacker, 2016; Lindgren, 2018) and this ability improves substantially with increasing age. At age 7, children include most such components in their narratives (see Figure 4). Comprehension of these aspects was not assessed, but based on the children's production, the children were able to comprehend the actions shown in the pictures correctly (and tell about them too). Second, the extent to which children spontaneously included information about characters' goals and internal states (which needed to be inferred from the pictures) in their narratives did not improve through the ages studied. However, the analysis of the children's performance on the comprehension questions showed that children are able to draw inferences about these aspects from the pictures and verbalize them when explicitly probed. This ability was partially present already at 4 but developed extensively up to age 7. The study thus shows that the ability to tell stories that include aspects of narrative macrostructure shown in pictures has been acquired by age 7, whereas this is only the case for macrostructural components requiring inferencing when children are explicitly asked.

To conclude, the results of the present study do not only point to consistent and large differences in children's ability to comprehend and produce narrative structures from age 4 to 7, but also to the influence of the stimulus materials and the specific way in which narrative macrostructure is scored. For this reason, in order to be able to compare results for the development of macrostructure, it is important that future studies use similar stimuli and scoring protocols.

Acknowledgements

The author is grateful to Natalia Gagarina and the other members of the Leibniz-ZAS Research Area 'Language Development and Multilingualism', an anonymous reviewer and the editors for valuable comments on earlier versions of this article; to Ute Bohnacker and the BiLI-TAS team at Uppsala University for extensive discussions about scoring comprehension and production of macrostructure in MAIN; to the children and their parents; and to the preschool teachers who helped with recruitment.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Notes

1. All the results reported below are from studies in which narratives were elicited in the telling mode using MAIN *Baby Birds*/*Baby Goats*, the task type and the stories used in the current study.
2. In their longitudinal study, Blom and Boerma (2016) report comprehension scores from questions asked after the child listened to a model story that was *Cat*/*Dog* from the MAIN; these scores are not comparable to the results from the current study. It should be noted, though, that Blom and Boerma (2016) found no increase in the typically developing children's comprehension scores from age 5;9 to 6;9.
3. Fifteen of the participants were part of the 4-year-old group in a recent cross-sectional study of Swedish-speaking mono- and bilingual children's narrative ability (Lindgren, 2018).
4. All parents had completed secondary education; most also had at least some tertiary education.
5. The reason for this procedure was that 15 of the children were (as 4-year-olds) part of a larger cross-sectional study, in which performance on different narrative tasks was compared (Lindgren, 2018).
6. The two groups also did not differ significantly in their scores on production of verbs and nouns, $t(15) = 2.00$, $p = .064$ (*Baby Birds* group $M = 46.3$ points, *Baby Goats* group $M = 42.4$ points), two subparts of the Swedish version of the Cross-linguistic Lexical Task (Ringblom, Håkansson, & Lindgren, 2014), a vocabulary task developed for preschool children (see Haman, Łuniewska, & Pomiechowska, 2015).
7. This category thus included cases in which only a goal was produced (6 cases in the data), as such episodes do not contain any sequence of components, but only a single component.
8. Running a repeated-measures ANOVA on producing GAO or not produced the same result.

ORCID iD

Josefin Lindgren https://orcid.org/0000-0003-3995-8315

References

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.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove, UK: Psychology Press.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech and Hearing Research*, *35*, 119–129.

Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities*, *55*, 301–311.

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. *International Journal of Language & Communication Disorders*, *51*, 626–638.

Bohnacker, U. (2013, March). *Language impairment or typical language development? Developing methods for linguistic assessment of bilingual children in Sweden* (Project usually referred to as BiLI-TAS). Grant proposal to the Swedish Research Council (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*, 19–48.

Bohnacker, U., & Lindgren, J. (in press). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. Grohmann (Eds.), *LITMUS in action: Cross-comparison studies across Europe*. Amsterdam, The Netherlands: John Benjamins.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21.

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.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, *35*, 155–168.

Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: Rasskaz and Erzählung. *Applied Psycholinguistics*, *37*, 91–122.

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., 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, UK: Multilingual Matters.

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2018). The receptive–expressive gap in English narratives of Spanish–English bilingual children with and without language impairment. *Journal of Speech Language and Hearing Research*, *61*, 1381–1392.

Gillam, R. B., & Pearson, N. (2004). *Test of narrative language*. Austin, TX: Pro-Ed.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, England: Longman.

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, UK: Multilingual Matters.

Hayward, D., Schneider, P., & Gillam, R. B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research*, *55*, 54–72.

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.

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*, 177–191.

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, *37*, 123–144.

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: University of Washington.

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6*. Uppsala, Sweden: Acta Universitatis Upsaliensis.

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, *29*, 327–365.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, *37*, 69–89.

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press.

Nordqvist, Å. (2001). *Speech about speech: A developmental study on form and function of direct and indirect speech* (Doctoral thesis). University of Gothenburg, Gothenburg, Sweden.

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*. Advance online publication. doi:10.1080/13670050.2018.1434124

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*, 36–76.

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, UK: Multilingual Matters.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum Press.

Reuterskiöld, C., Hansson, K., & Sahlén, B. (2011). Narrative skills in Swedish children with language impairment. *Journal of Communication Disorders*, *44*, 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*, 83–93.

Ringblom, N., Håkansson, G., & Lindgren, J. (2014). *Cross-linguistic Lexical Task: Swedish version (CLT-SWE)*. Unpublished material.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, *37*, 49–67.

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*, 224–238.

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., & Day, D. (1993). On the development of narrative structure in child L1 and adult L2 acquisition. *Applied Psycholinguistics*, *14*, 135–158.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*, 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*, 133–170.

van den Broek, P., Kendeou, P., Kremer, K., Lynch, J. S., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Stahl, & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum.

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, The Netherlands: John Benjamins.

Westby, C. E. (2012). Assessing and remediating text comprehension problems. In A. G. Kamhi, & H. W. Catts (Eds.), *Language and reading disabilities* (3rd ed., pp. 163–225). Boston, MA: Pearson.