Macrostructural organization of adults' oral narrative texts[1]

Natalia Gagarina
Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

Ute Bohnacker
Uppsala University, Sweden

Josefin Lindgren
Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

This study investigates macrostructure in elicited narratives of 69 monolingual German-, Russian- and Swedish-speaking adults. Using the LITMUS-MAIN (Multilingual Assessment Instrument for Narratives), and its Baby Goats and Baby Birds stories, story structure and story complexity, concerning episodic organization, were examined across the 3 languages. As theoretical underpinnings, a multidimensional model of macrostructure was used. This model includes analyses of story structure (SS), in which a narrative merits a maximum score of 17, based on the occurrence of five types of macrostructural components (Internal states as initiating event and as reaction, Goal, Attempt and Outcome), and of story complexity (SC), which measures combinations of Goals, Attempts and Outcomes within one episode. The highest attainable complexity is the GAO-sequence, when a Goal, Attempt and Outcome are produced within the same episode. The results for SS were similar for German, Russian and Swedish, where adults included 11-12 components per story. A more detailed analysis of the individual components revealed striking similarities across the 3 languages, both for frequently used and seldom occurring components. SC did not differ significantly across languages nor across stories, whilst for SS, a slight difference between the two stories was found. We interpret this finding as story complexity (a qualitative measure of macrostructure) being of a more universal nature. Furthermore, our results indicate that caution is warranted when conclusions about children's narrative skills are to be drawn on the basis of the MAIN Baby Goats and Baby Birds stories.

1   Introduction

Oral and written texts form an essential part of our human life. Without being able to produce and comprehend coherent and cohesive texts, communication would fail. In and outside the home, in school and later on in professional life, texts play a crucial role (Barton, 2007; Bliss, McCabe, & Miranda, 1998; Janks, 2013; Westby, 2012). Considering the importance of texts throughout the lifespan, surprisingly little research on the global organization (or macrostructure) of oral texts has been done. While microstructure (e.g. text cohesion) has been at the center of attention for many years and has been examined from different perspectives by employing various theoretical approaches (Halliday & Hasan, 1976; Restrepo & Kruth, 2000; Thordardottir, Weismer, & Smith, 1997), macrostructure has received considerably less attention in both theoretical linguistics and empirical education research. However, recently a common interest in the examination of macrostructure within the field of language acquisition has brought together an interdisciplinary group of researchers who aimed to fill this gap. Within the COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment" (2009–2013), this group of researchers first studied the existing instruments for assessing the narrative skills of children across languages, such as the Bus Story (Renfrew, 1969), the Test of Narrative Language (TNL; Gillam & Pearson, 2004), the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2005), and the HAVAS (Reich & Roth, 2004), and examined whether these instruments are cross-culturally robust and would be suitable for different populations of multilingual children. They also examined in detail different theories and models of macrostructure, such as the story grammar framework (Stein & Glenn, 1979), that had previously been used for evaluating the organization of elicited oral texts on a global level. Since they were not able to find an instrument that would satisfy their criteria, the decision was made to create a new instrument. This instrument should include clear and systematic pictorial realization of story components based on a well-considered theoretical model of macrostructure and consist of several stories that are parallel in pictorial content. The new instrument should also be suitable for the cultures and populations that the participants of the COST Action were working with in Europe, Africa and beyond.

A multidimensional model of the global organization of story structure was developed, and sketches were made of different story plots. The main components of this multidimensional model were visualized by a Lithuanian professional artist, Loreta Valantiejiene. The entire process from the initial drafts of the stories up to the final versions of the picture sequences took more than four years. During these years, an interdisciplinary team of researchers from about 20 countries,

including a core group of representatives from eight countries, Cyprus, Finland, Germany, Israel, Lithuania, Russia, South Africa and Sweden, discussed and further developed the content of the stories, the flow of plot, and its visual realizations. The artist painted and repainted more than 200 versions of individual pictures from the picture sequences. Finally, the *Multilingual Assessment Instrument for Narratives* (MAIN), a new instrument for the assessment of the global organization of narrative texts (as well as their microstructural organization) had been created (Gagarina et al., 2012, 2015). This instrument consists of four parallel picture stories (Cat, Dog, Baby Birds, Baby Goats) with six pictures each. All four stories all have similar story structure and story complexity, following the multidimensional model of the global organization of stories. There is parallelism in the number of episodes (each story consists of three episodes) and in the composition of episodes. Special efforts were made to achieve parallelism in the choice of story characters (protagonists) and actions, including their cognitive appropriateness for young children, and how the protagonists' facial expressions and bodily movements are visualized. The order in which the protagonists appear in the picture was parallelized as well. Efforts were also made to give the backgrounds, objects and smaller details in the pictures a clean and unified look, so that they can be perceived clearly by children of different backgrounds and cultures.

2   Macrostructure and macrostructure evaluation in MAIN

Macrostructure is the global organization of a text. In narratives, macrostructure can be captured, for instance, via the *story grammar model* (G. Mandler, 1979; Stein & Glenn, 1979).[2] According to Stein and Glenn (1979), story grammar includes the setting (which introduces the characters and describes the non-linguistic context of a story in time and space) and one or several episodes. An episode in its turn consists of an initiating event, internal response, internal plan, attempt, direct consequence and a reaction (Stein & Glenn, 1979). The number and the naming of these components might differ between narrative approaches (e.g. Labov, 1972; Labov & Waletzky, 1967; J. M. Mandler & Johnson, 1977; Peterson & McCabe, 1983; Stein & Glenn, 1979; Stein & Policastro, 1984). Not all components need to be overtly realized in the production of narrative text; thus, episodes might be not exhaustively realized when some of the components are omitted. The combination of verbalized components of an episode leads to more or less complex sequences (i.e. more or less complete episodes). In order to capture different levels of complexity, Westby (2012) suggested a binary decision

---

[2]   Various different systems of analyzing the global organization of a text and of evaluating its quality exist, but this is not the topic of the present study.

tree, which distinguishes between: i) no sequence of events, ii) action sequence, i.e. Attempt-Outcome (AO) sequence with no Goal (G) statement, iii) abbreviated episode, with a G statement and either A or O (GA, GO), and iv) complete episode with all three GAO components. According to a number of researchers, the GAO represents the highest level of story complexity and is an indicator of the ability to produce a coherent story (Stein & Policastro, 1984; Trabasso & Nickels, 1992; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992; Westby, 2012).

Our own instrument, the MAIN, not only builds upon these ideas of story grammar components, but extends them and suggests a new system of organization of text components on the macrostructural level (see Gagarina et al., 2012, p. 20). We suggest that narrative texts contain complete units (i.e. episodes) which consist of five components: two components, an initiating one and a reaction, frame the three components that form the core content of the episode. These three core components describe (a) the objective of the protagonists' action, called Goal, (b) the protagonists' action itself, called Attempt, and (c) the accomplishment (or not) of this action, called Outcome. The two framing elements are, firstly, the internal state of the protagonist at the outset of the episode, e.g. an emotional or cognitive state which triggers or initiates the action itself, i.e. leads to the goal, and secondly, the internal state of the protagonist as a reaction to the outcome of the action. These framing elements are referred to as internal state terms (here abbreviated as IS) in MAIN. This model generates a systematically structured episode consisting of five components. In order to give a child more possibilities to verbally realize these, in MAIN, episodes are systematically visualized three times per story in parallel picture sequences with different animate protagonists. While actions of the protagonists are explicitly depicted, goals and internal states of the story characters must be inferred from the pictures.

In order for a narrative text to be complete, an appropriate beginning, or setting, is needed as well. In MAIN, the realization of a setting statement mentioning time and place is an integral part of the evaluation of macrostructure. Such setting statements are for example, *once upon a time* (time) or *on a meadow* (place).

Realizations of the narrative macrostructure can be elicited in MAIN with four parallel tasks via comparable pictorial stimuli. Macrostructure evaluation in MAIN is multidimensional: It can be assessed quantitatively, via the sum of episode components in the whole narrative text (i.e., story structure, SS), and qualitatively via the combination of episode components (i.e., story complexity, SC). For story structure, the presence of the following components is totalled: Setting information (max 2 points, 1 for time, 1 for place) and the total number of components per episode, which is five: Goals (G), Attempts (A), Outcomes (O), ISs produced as an initiating event, and ISs produced as a reaction. Each story

contains three episodes, thus the maximum number of points is 17 (2 for setting, 9 for each of three Gs, As, and Os, and 6 points for ISs as either initiating events or reactions) per narrative. Repetitions or multiple verbalizations of the same component are excluded from this calculation. Table 1 shows the setting and one full episode (Episode 1) for the MAIN Baby Goats story.

Table 1. 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.

All three episodes of the Baby Goats story are shown in the pictures in Figure 1, and in an idealized script of that story, where every single one of the possible components is overtly realized and highlighted (see Table 2).

In Table 2, 17 components are marked, showing the maximum score of the MAIN story structure score (17 points). The story structure score (SS) is the quantitative evaluation of macrostructure. The occurrence of three core components (goal, attempt, outcome, GAO) in each episode reflects the highest story complexity (SC) possible, which is the qualitative evaluation of macrostructure in MAIN. Since there are three episodes per story, a GAO can occur three times per story.

Figure 1. Small-scale black-and-white copy of the Baby Goats pictures.

Table 2. Story script with macrostructural components marked, Baby Goats.

	Component	SS	SC
One day there was a mother goat who *saw* that her baby goat had fallen into the water and that it was *scared*.	Ep1 IS as IE	1	
She jumped into the water	Ep1 A	1	**Ep1**
because she wanted to save it.	Ep1 G	1	**GAO**
A *hungry* fox *saw* that the mother goat was in the water and growled: "Mmm, nice, what do I see here on the grass?".	Ep1 IS as IE	1	
The mother goat pushed the baby goat out of the water, but she did not see the fox.	Ep1 O	1	
She was **glad** that her baby did not drown.	Ep1 IS as R	1	
Meanwhile the mean fox jumped forward	Ep2 A	1	**Ep2**
because he wanted to catch the other baby goat.	Ep2 G	1	**GAO**
He grabbed the baby goat.	Ep2 O	1	
The baby goat was **scared**.	Ep2 IS as R	1	
A brave bird that was flying by *saw* that the baby goat was in great danger.	Ep3 IS as IE	1	
He decided to stop the fox and save the baby goat.	Ep3 G	1	**Ep3**
The bird said to the fox: "Leave the baby goat alone". And then he flew down and bit the fox's tail.	Ep3 A	1	**GAO**
The fox let go of the baby goat and the bird chased him away.	Ep1 O	1	
The bird was very **happy** that he could save the baby goat, and the fox was still **hungry**.	Ep1 IS as R	1	
	Total SS	17	

Note. The component types are highlighted as follows: *internal state as initiating event,* goals, attempts, outcomes, **internal state as reaction**. Ep1 = Episode 1, Ep2 = Episode 2, Ep3 = Episode 3, IS = internal state, IE = initiating event, R = reaction, SS = story structure, SC = story complexity.

3   Aims, research questions and hypothesis

The present study investigates narrative macrostructure in monolingual adults speaking German, Russian and Swedish. It applies the evaluation system of MAIN and thereby provides much needed adult benchmark data for Story structure (SS) and Story complexity (SC). Since in previous work (e.g. Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Bohnacker, 2016; Gagarina et al., 2015; Lindgren, 2018; Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, & Haman, 2018), narrative macrostructure has been shown to develop with age but to be less dependent on mono-/bilingualism, our adult data can serve as a yardstick for mono- and bilingual children's narrative skills. The adult narrators (see below) either told the MAIN Baby Birds or the Baby Goats story.

Our first research question is: What is the adult story structure and story complexity for the three different languages and across stories? Based on previous results that show similarities in the story structure and story complexity across the two languages of bilingual children and across various age-matched monolingual

(and bilingual) children (Bohnacker, 2016; Gagarina et al., 2015; Kunnari, Välimaa, & Laukkanen-Nevala, 2016), we predict that narrative macrostructure will not significantly differ across the three languages in adults either. Neither do we expect any significant differences across the two stories (Baby Birds, Baby Goats), because they were constructed to be parallel in terms of macrostructure. It could be the case however that we find more variation between stories in SS than in SC.

The second research question concerns the episodic structure of oral texts: What is the frequency of the episode components and what types of components within the three different episodes are verbalized by the adult speakers of the three languages? Here, we predict differences in verbal realization between different episodes/stories, since they depict different situations, such as a baby goat sinking in an expanse of water, or baby birds with wide open beaks in a nest, being hungry. There is a difference between, for example, saving someone's life and feeding someone, and thus we expect that the participants may realize the components of these episodes differently. This prediction is grounded in previous findings for Russian – spoken by children in 6 different European countries (Gagarina et al., in press), and for monolingual and bilingual Swedish-speaking children (Lindgren, 2018; 2019), where differences in the production of the three episodes within and across the Baby Birds and Baby Goats stories emerged. Although the episodes were constructed to be identical, they appear to be different regarding the *strength* of the initiating event, i.e. (a baby goat) drowning vs. (baby birds) being hungry. This variation in the *real-world* situation might lead to variability in the verbalization of story components. So, we expect differences across the episodes/stories in the expression of internal states as initiating events, goals, attempts, outcomes, and internal states as reactions.

4   Method
4.1   *Participants*

The participants were 69 adult speakers of German (N = 30, M = 29.5 years, 16 women), Swedish (N = 19, M = 28.1 years, 11 women), and Russian (N = 20, M = 26.5 years, 13 women) aged 19–41 years. A Welch one-way analysis of means (not assuming equal variance in the groups) showed that there was no significant difference in age between the groups ($F(2, 29.77) = 3.13$, $p = .06$). All participants had grown up monolingually, although many of them had learned additional languages later in life, such as foreign languages in school. All participants had had typical language development and had at least finished secondary school, and most of them were either attending or had completed tertiary education.

4.2   *Narrative elicitation, transcription and coding*

A native speaker of the respective languages administered Baby Birds/Baby Goats from the MAIN (Gagarina et al., 2012; 2015) to each participant in a quiet room. As mentioned in Section 1, Baby Birds and Baby Goats are picture sequences with six pictures depicting three-episode stories.[3] The episodes contain carefully constructed goal-attempt-outcome sequences for specific characters. Baby Birds and Baby Goats are parallel in terms of length and story grammar components, and both contain five story characters. Each participant told one story, either Baby Birds or Baby Goats, and for each language, the number of participants telling these stories was counterbalanced. In total, 35 participants told Baby Birds, and 34 told Baby Goats. The German participants told Baby Birds/Baby Goats after having listened to another ('model') story (Cat/Dog) and answering comprehension questions about it. The Swedish participants had first told the Cat/Dog story and answered the comprehension questions about it before telling Baby Birds/Baby Goats, as part of a larger test battery (for details, see Lindgren, 2018, Chapter 3). The Russian participants told Baby Birds/Baby Goats with no preceding task. For all 69 participants then, the Baby Birds/Baby Goats was a story generation task; they all told the stories from the MAIN picture stimuli only (i.e. none had listened to the story before).

The general procedure for the MAIN was followed. Before and during the story telling, the pictures were only visible to the participant, not to the experimenter. The experimenter only gave general feedback signals (*mm*, *aha*) and in rare cases gave some prompting (e.g. *and then*?). For comparability, exactly the same stimulus materials and elicitation procedures were used with all three groups of speakers.

The narratives were transcribed in CHAT format (MacWhinney, 2000) by trained transcribers who were native speakers of the language. The coding of macrostructure (see below) was carried out according to the MAIN manual by native speakers of the language. Part of the data was coded by a second coder who was a native or fluent speaker of the language. For Swedish, five narratives (26% of the data) were recoded and the interrater agreement was 93%. For German, eight narratives (27% of the data) were recoded and the interrater agreement was 92%. For Russian, six narratives (30%) were recoded and the interrater agreement was 97%.[4]

---

[3]   Additionally, the MAIN includes ten comprehension questions per story. These questions target comprehension of characters' goals and internal states.

[4]   Note that for Russian, interrater agreement was calculated based on the remaining disagreements after discussions between the two scorers, whereas for German and Swedish, interrater agreement was based on differences in scoring between the two scorers without discussion.

All narratives were coded for the absence or presence of each of the different *types of macrostructural components* (setting + IS as initiating event, goal, attempt, outcome, IS as reaction for each of the three episodes in the story), and then received a total *story structure* (SS) score (maximum score = 17 points). For *story complexity* (SC), we focused on the production of sequences of the core macrostructural components (goal, attempt, outcome), and coded each episode in the narratives according to the type of sequence of macrostructural components that were produced as either no sequence (only *one* out of the components goal, attempt and outcome was produced for that episode), AO-sequence (attempt + outcome), GA/GO-sequence (goal + attempt or goal + outcome), or GAO-sequence (goal + attempt + outcome). Each narrative was also coded for the highest level of macrostructural complexity reached.

5   Results

We first analyzed the story structure scores, focusing on differences between languages and stories (Section 5.1). Next, we took a closer look at the frequency with which the different components were included in the narratives, focusing on the types of components in the different episodes (Section 5.2). Finally, we analyzed macrostructural complexity for the languages, stories and episodes (Section 5.3).

5.1   Story structure score

In Table 3, story structure scores for the three languages are shown. Scores for the three languages are relatively similar (mean 11 or 12 points), with similar variance in scores, both in terms of SDs and score ranges. Although there are participants in each language who score close to the maximum score (17 points), this is far from the case for all adults; the mean scores are only at 65%–70% of the maximum, and some adults scored relatively low.

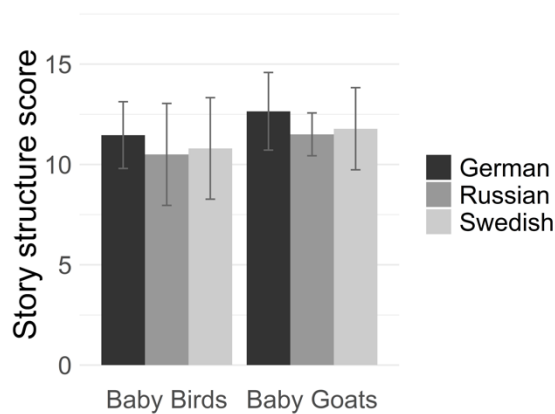**Table 3.** Story structure score, by language (Maximum = 17 points)

	Mean	SD	Range
German	12.1	1.9	9 – 16
Russian	10.9	2.1	6 – 15
Swedish	11.3	2.3	7 – 15

In Table 4, an overview of the scores is given for the two stories, Baby Birds and Baby Goats.

Table 4. Story structure score, by story (Maximum = 17 points).

|	Mean	SD	Range
Baby Birds	10.9	2.2	6 – 15
Baby Goats	12.2	1.8	8 – 16

A Language x Story (3 x 2) factorial ANOVA showed that there was no significant effect of language ($F(2, 65) = 1.816$, $p = .17$), but that scores on Baby Goats were significantly higher than scores on Baby Birds ($F(1, 65) = 4.817$, $p = .03$).[5] The consistent, but relatively small, effect of story across languages can be seen in Figure 2.

Figure 2. Mean story structure scores, by language and story. Error bars show ±1 SD. (Maximum = 17 points).
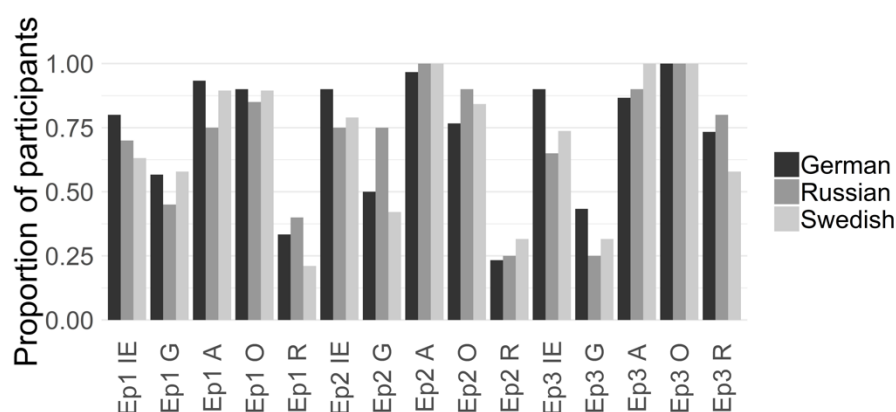
5.2 *Types of macrostructural components*

Next, we took a closer look at how often the different components (except for settings) were produced by the participants. Figure 3 gives an overview of the proportion of the participants, irrespective of language, who produced the different components.

---

[5]   Adding the interaction effect (Age x Language) did not improve model fit ($p = .98$).
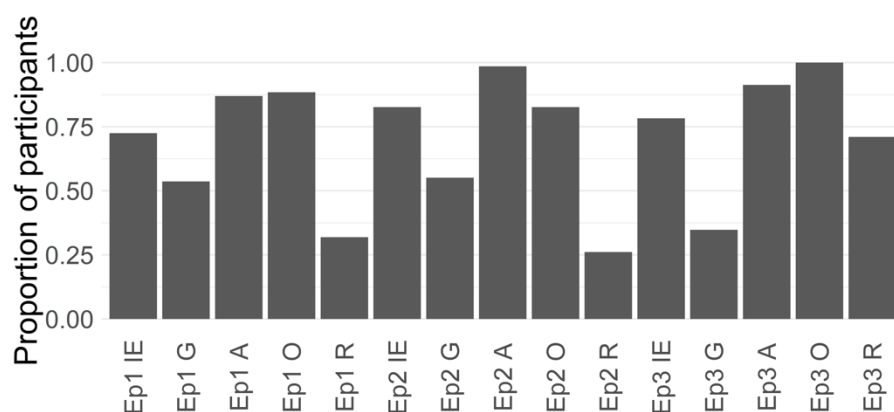
Figure 3. Macrostructural components, proportion of all participants who included the component in their narrative (N = 69).

Figure 3 shows that not all components are produced equally often by the participants: attempts and outcomes are included by the large majority of the participants (at least 80%), irrespective of episode, and in the case of Attempt in Episode 2 and Outcome in Episode 3, by close to or 100% of the participants. The component seen as the most central for a well-formed narrative, the goal, was produced much less frequently, at around 50% in Episodes 1 and 2, and in only around 35% of the narratives for Episode 3. All goals are thus not equally often included in narratives. In contrast, ISs as initiating events were very frequent in the participants' narratives, with around 75% or more participants producing them for each episode. For ISs as Reaction, there was a notable difference between Episodes 1 and 2, on the one hand, in which this component was very infrequently produced (only by around 30% of the participants) and Episode 3, on the other, in which this component was produced by close to 75% of the participants. Patterns were strikingly similar in the three languages, as shown in Figure 4.

Figure 4. Macrostructural components, proportion of participants who included the component in their narrative, by language (N = 69).

5.3   Story Complexity (SC)

For story complexity, we first analyzed the proportion of participants who reached each of the different levels of macrostructural complexity at least once in their narratives. The results are shown in Figure 5. We found no significant difference between the languages ($\chi^2(4, N = 69) = 0.479$, $p = .976$). In all three languages, the majority of the speakers (German: 77%, Russian: 70%, Swedish: 74%) used at least one GAO-sequence, and thus reached the highest level of complexity in their narrative. A smaller group did not produce anything more complex than AO-sequences (German: 23%, Russian: 10%, Swedish: 26%), and in Russian, four speakers (20%) produced a GA/GO-sequence as their most complex episodic structure.
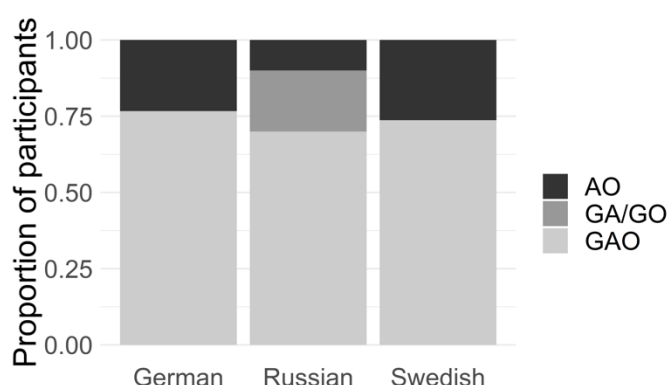

Figure 5. Proportion of the participants who reached the different complexity levels, by language (N = 69).

Next, we proceeded to look instead at all episodes within all the narratives. Although at least 70% of the adults produced at least one GAO, when all episodes in the narratives (N = 207) were considered, the picture looks somewhat different in terms of complexity. Less than half of all episodes (German: 40%, Russian: 40%, Swedish: 40%) contained a GAO-sequence. GA/GO-sequences were rare (German: 9%, Russian: 7%, Swedish: 4%). Approximately half of the episodes were AO-sequences (German: 42%, Russian: 43%, Swedish: 49%). A smaller proportion of the episodes (German: 9%, Russian: 10%, Swedish: 7%) did not contain any sequence, i.e. at least two, of the core macrostructural components (goal, attempt, outcome). Again, the patterns in the three languages were the virtually identical ($\chi^2(6, N = 207) = 2.192$, $p = .901$), as shown in Figure 6.
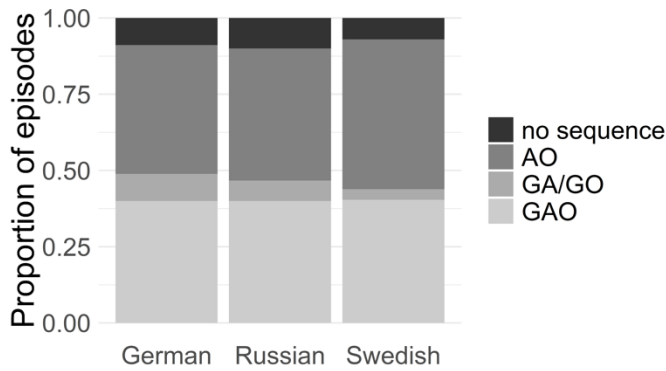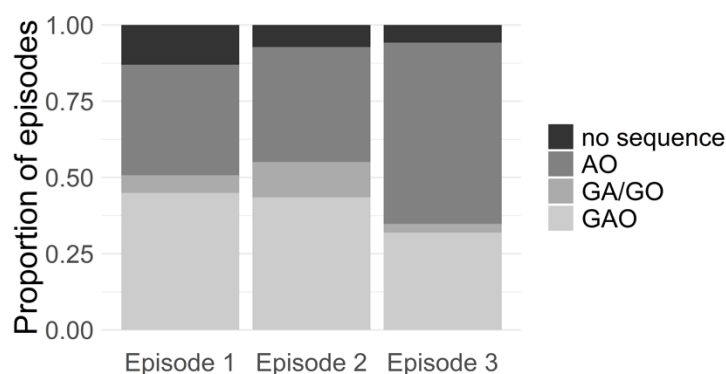
Figure 6. Types of sequences produced in the episodes (N = 207), by language.

Despite the fact that the Baby Goats narratives tended to receive a higher story structure score than Baby Birds (recall Section 5.1), the patterns of macrostructural complexity were very similar for the two stories (χ^2(3, N = 207) = 1.229, p = .746), as shown in Figure 7.
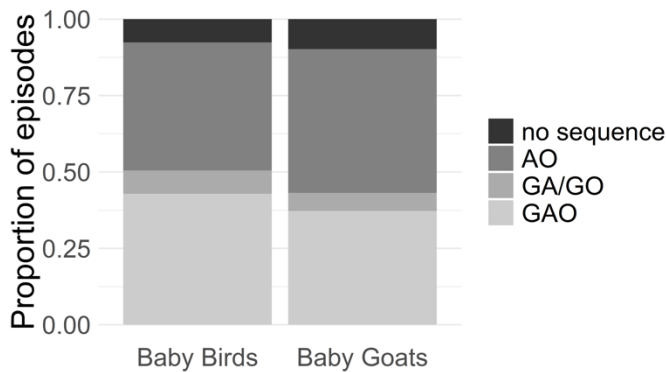
 Figure 7. Types of sequences produced in the episodes (N = 207), by story.

Results for macrostructural complexity in the three different episodes of each story are shown in Figure 8. Complexity differed significantly between the three episodes (χ^2(6, N = 207) = 13.331, p = .038). AO-sequences were more common in Episode 3 (59%) compared to the other two episodes (36% and 38% for Episode 1 and 2, respectively). GA/GO-sequences were somewhat more frequent in Episode 2 (12%) compared with Episode 1 (6%) and Episode 3 (3%). Interestingly, Episode 1 had the highest proportion of both GAO-sequences (45%) and no sequence (13%), compared with the other episodes; Episode 2 had a higher proportion of GAO-sequences (44%) compared with Episode 3 (32%), whereas the proportions of no sequence were similar in Episode 2 (7%) and Episode 3 (6%). Complexity was thus somewhat lower in Episode 3.

Figure 8. Types of sequences produced in the episodes (N = 207), by episode.

6 Discussion and conclusion

The present study has examined the macrostructure of elicited narratives for three groups of adults, speaking German, Russian or Swedish. The narratives were collected using the Baby Birds and Baby Goats stories from the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012; 2015). For the evaluation of story structure and story complexity, a multi-dimensional model of macrostructure was used. The aim was to determine adult story structure and story complexity in three different languages and to scrutinize episodic complexity in depth. First, story structure scores across languages and stories were analyzed, second, the frequency of the use of different components was investigated, and finally, story complexity was examined. The results of our study cannot be compared to any other study, as no comparable analyses of the macrostructure in adults have been carried out (but cf. Stein & Glenn 1979). So, we will give an overview of the results and discuss them in relation to the cognitive and *real-life* context and language acquisition.

First, the results show that adult story structure and story complexity manifest in the three languages in very similar ways. Interestingly, the mean scores of our 69 adult speakers were still relatively far from the maximum of 17 points for story structure on MAIN. This is an important result as it shows that a *normal,* typical adult produces on average 11 to 12 components per story and not all 17 components (as in the idealized script in Table 2), even though some of our speakers did express 15 or 16 components. Since we are talking about the *target* here, i.e. what we expect children to reach in an ideal case, this finding should be considered when evaluating the performance of children who are still developing their narrative skills. Our findings for adults suggest that if a child reaches 11 or 12 points out of 17, this can be considered adultlike; the child has then acquired

appropriate use of story structure. Thus, our results provide us with a benchmark for the minimalist acquisition target, which is not 17, but 11 points.

We also found for all three languages (German, Russian and Swedish) that the adult story structure scores for Baby Goats were slightly but significantly higher than for Baby Birds. This is in line with similar results from a longitudinal study of Swedish monolingual children aged 4–7 by Lindgren (2019), and has also been found for story comprehension (Bohnacker & Lindgren, in press; Lindgren, 2018; 2019). The significant difference found between the two stories, Baby Goats and Baby Birds, in the present study indicates that researchers should be careful when comparing Baby Goats (M = 12.2) and Baby Birds (M = 10.9) stories in children, because lower performance in Baby Birds might be an artefact of this story. On the other hand, 11 vs 12 components realized is not a very large difference, so this finding also shows that a significant difference should be treated critically.

Second, a closer look at the different macrostructural components showed that production was strikingly similar in the three languages, both for the frequently used and seldom occurring components. This is good evidence for the pictorial stimuli being cross-linguistically and cross-culturally robust, at least for three groups of monolingual adults (all Europeans, and all literate). Some of the components were produced very frequently, particularly attempts and outcomes (at least 80%), while others, especially goals, were produced less often (around 50% goals in Episodes 1 and 2, and around 35% in Episode 3). This finding gives rise to an intriguing question about the goal component of stories: if this very component is central in the evaluation of story structure, how can it be that adults overtly produce it in only 35% of cases when they narrate an episode? This finding invites us to reconsider the 'technical' part of the evaluation of story structure and story complexity in which goals are central elements. In contrast to goals, ISs as initiating events were common in all episodes, and ISs as reactions were very frequently produced in Episode 3 (by close to 75% of the participants), but not in Episodes 1 and 2. So, the following picture arises from these results: ISs as initiating events, attempts and outcomes are the most stable components of story structure in the sense that they are very frequently produced in all episodes, stories and languages. The other elements are more vulnerable; their use depends on the type of episode, story, etc. Having this in mind, the question arises whether the evaluation of story structure should be reconsidered in such a way that it primarily includes the most stable, less vulnerable components, such that it is not affected by 'additional' factors which might blur or impede adequate evaluation. The question arises whether such a robust representation can be achieved more generally.

Third, as predicted, story complexity was not found to differ significantly, neither between languages nor between stories (Baby Goats vs. Baby Birds). This

finding suggests that story complexity (i.e. the qualitative evaluation of macrostructure) may be of a more universal nature than story structure. We explored story complexity by analyzing the proportion of speakers in a group that produced at least one GAO in their narrative (as suggested in Gagarina 2016), and this proportion was generally high (70%–77%). The ability to produce (at least) one full GAO-sequence per narrative may thus be a promising qualitative measure of whether macrostructure is mastered or not.

We also tackled story complexity from another angle, evaluating the number of episodes in which full GAO-sequences were realized by the participants. In all three languages, only 40% of the episodes were realized as full GAO-sequences. This may again be due to real-life differences between episodes, which are denoted in the pictures. It may not be as necessary to express the Goal 'the mother goat wants to save her baby' in a situation where someone is shown to be drowning, and a quick reaction is necessary. A natural way to capture this situation in storytelling (and also frequently attested in our adult data) would be to say that the mother goat ran into the water and saved her baby goat (or pushed it out of the water), thus only verbalizing the attempt and outcome.

We recommend that future studies of MAIN with children take into account our finding that there is a slight but significant difference in how adults realize the macrostructure in the two stories (Baby Birds, Baby Goats). It could be the case that results are more alike for the other two MAIN stories, Cat and Dog, because episodic realization of story content is more similar in Cat and Dog than in Baby Birds and Baby Goats; and the Cat and Dog stories are more uniform, parallel and comparable as far the main protagonists, their actions as well as other factors are concerned.

As an epilogue to this study, we should mention that alternative models of a full episode might be employed and may give different results. Note that 'only' 70%–77% of our adult speakers produced a full GAO-sequence in their narrative, the type of sequence which is commonly regarded as a full episode (Stein & Policastro, 1984; Trabasso & Nickels, 1992; Trabasso et al., 1992; Westby, 2012). We also found that Goals, commonly seen as the most central part of an episode, were not always overtly realized, and yet our adults cannot be considered to be 'bad' or atypical storytellers, but normal, competent adult speakers. Since the Goal appeared not to be the most stable component adults verbalize in their oral narratives, the question arises whether the Goal should be used as the cornerstone measure for the evaluation of full episodic complexity. It might perhaps be more in line with real-life storytelling to rely on the most robust components of episodic structure. In addition to the 'strict' way of defining a full episode as GAO, as was done in the present study, we therefore also ran two alternative analyses of episodic complexity on our adult data. These will be elaborated on in a forthcoming study. As goals were less frequently included than ISs as initiating

events (see Section 5.2), and since ISs as initiating events also can be the starting point of an episode, we recoded the episodes according to whether they included an IS as IE in addition to an attempt and an outcome, forming an 'alternative GAO', an IAO-sequence (IS as initiating event + attempt + outcome). Next, we recoded the data for initiating *events* (in addition to *internal states as initiating events*), to create yet another type of evaluation of episodic complexity, which can be called another 'alternative GAO-sequence', the EAO-sequence (IE + A + O). This way of operationalizing complexity follows earlier suggestions by Stein & Glenn (1979). The two 'alternative GAO' evaluation models (IAO, EAO) were used to analyze episodic complexity. Finally, we combined all three complex sequences (GAO, IAO, EAO) into a 'modified GAO' – the empirical results for these alternative models of story complexity will soon see the light of the day.

7   Acknowledgements

8   References

Barton, D. (2007). *Literacy: An introduction to the ecology of written language* (2nd ed). Malden, MA: Blackwell Pub.

Bliss, L. S., McCabe, A., & Miranda, A. E. (1998). Narrative Assessment Profile: Discourse analysis for school-age children. *Journal of Communication Disorders*, *31*(4), 347–362.

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.

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, *37*(1), 19–48.

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

Gagarina, N., Fichman, S., Galkina, E., Protassova, E., Ringblom, N., & Rodina, Y. (in press). Macrostructure in Russian oral texts across Europe. In S. Armon-Lotem & K. K. Grohmann (Eds.), *LITMUS in Action: Cross-Comparison Studies across Europe*. Amsterdam: John Benjamins.

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*.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of Narrative Abilities in Bilingual Children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children disentangling bilingualism from language impairment* (pp. 243–269). Bristol: Multilingual Matters.

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: Longman.

Janks, H. (2013). Critical literacy in teaching and research. *Education Inquiry*, *4*(2), 225–242.

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.

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington.

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

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (Third Edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mandler, G. (1979). Organization and repetition: Organization principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research* (pp. 293–327). Hillsdale, NJ: Erlbaum.

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

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*, 1–25.

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

Renfrew, C. E. (1969). *Bus Story Test: A Test of Narrative Speech*. Bicester, Oxon: Winslow Press.

Restrepo, M. A., & Kruth, K. (2000). Grammatical characteristics of a Spanish-English bilingual child with Specific Language Impairment. *Communication Disorders Quarterly*, *21*(2), 66–76.

Schneider, P., Dubé, R. V., & Hayward, D. (2005). The Edmonton Narrative Norms Instrument. Retrieved November 17, 2015, from University of Alberta Faculty of Rehabilitation Medicine website website: http://www.rehabresearch.ualberta.ca/enni

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.

Thordardottir, E., Weismer, S., & Smith, M. (1997). Vocabulary learning in bilingual and monolingual clinical intervention. *Child Language Teaching and Therapy*, *13*, 215–222.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*(3), 249–275.

Trabasso, T., Stein, N. L., Rodkin, P. C., Munger, M. P., & Baughn, C. R. (1992). Knowledge of goals and plans in the on-line narration of events. *Cognitive Development*, *7*(2), 133–170.

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