How oral texts are organized in monolingual and heritage Russian

Gagarina, Natalia

Gagarina , N , Fichman , S , Galkina , E , Protassova , E , Ringblom , N & Rodina , Y 2021 , How oral texts are organized in monolingual and heritage Russian . in S Armon-Lotem & K K Grohmann (eds) , Language Impairment in Multilingual Settings : LITMUS in action across Europe . Trends in Language Acquisition Research , no. 29 , John Benjamins , Amsterdam , þÿp p . 4 8 7 5 . h t t p s : / / d o i . o r g / 1 0 . 1 0 7 5 / t i l a r . 2 9 . 0 2 g a g

acceptedVersion

How oral texts are organized in monolingual and heritage Russian. Evidence from six countries

Natalia Gagarina, Sveta Fichman, Elena Galkina, Ekaterina Protassova, Natalia Ringblom and Yulia Rodina
Leibniz-Center for General Linguistics (ZAS) / Bar Ilan University / Russian Academy of Sciences / University of Helsinki / Stockholm University / UiT The Arctic University of Norway

The present study compares narrative macrostructure, measured as Story Structure (SS) and Story Complexity (SC), in bilinguals speaking Russian as their home/heritage language (L1) and exposed to different societal languages (L2), while focusing on the effects of different L2s, bilingualism, and episodic struc- ture. The *Multilingual Assessment Instrument for Narratives* (LITMUS-MAIN) was used to elicit narratives in L1/Russian from 162 L2 Finnish, German, Hebrew, Norwegian, or Swedish bilinguals (4- and 6-year-olds) and 21 monolingual Russian children (4-year-olds). Age-matched bilinguals showed similarity in SS (except for children speaking L2 German or Hebrew) and SC. Monolinguals (age range 50–59 months) outperformed younger bilinguals (age range 48–59 months) in SS and SC but performed similarly to older bilinguals (age range 66–83 months). Fine-grained analysis revealed that a well-formed episode might include an Attempt-Outcome sequence combined with Internal States (and not only Goal-Attempt-Outcome) and that children are sensitive to events depicted in each episode. The findings show some evidence for the universality of macrostructure and provide insight into macrostructural knowledge at the episode level. The results are discussed within the theoretical model of multidimensional text organization.

Introduction

The Russian-speaking population living outside Russia is about 167 million people (Aleshkovski et al., 2018; Arefiev, 2017). This population includes people of different ethnicities and nationalities using Russian as their language of communication on an everyday basis, who organize communities, kindergartens, or schools and who transfer Russian to their children and grandchildren. For example, in Finland, Germany, Israel, Norway, and Sweden, Russian-speaking populations are well- established communities. In Finland, by the end 2019, 81,606 inhabitants declared Russian to be their mother tongue, excluding Finnish-Ingrian returnees, Russian- speaking immigrants of previous waves (the so-called 'Old Russians' who settled down in Finland in the 17th century and later and whose families maintained the Russian language), and children from binational marriages (Statistics Finland, 2020). In Germany, different publications indicate varying numbers of Russian- speaking population (including the so-called Russian Germans) with a maximum of over six million (Arefiev, 2017; cf. Report of the Russian Ministry of Foreign Affairs). In Israel, two immigration waves brought over 1 million immigrants from the former Soviet Union, resulting in 20% of the country's population (Amit, 2010; Remennick, 2003). Proportionally, fewer Russian-speaking immigrants are registered in Norway; their number includes approximately 22,000 Russian immigrants and Norwegian-born Russian speakers from mixed and non-mixed marriages (Statistics Norway, 2020). Finally, in Sweden, there are 29,000 Russian-speaking inhabitants (Parkvall, 2016, p. 276) and their number is increasing.

Russian immigrant groups are rather heterogeneous all over the world, but they are unified in their efforts to maintain their home language (heritage or first language, henceforth L1) and to transfer it to their children. Depending on the place of residence and the environment, there are different possibilities for language maintenance related to acquisition and learning intensity, regularity, and sustainability. While various background factors crucially affect the level of L1 proficiency in the production and perception of spoken and written Russian (Armon-Lotem et al., 2011; Gagarina et al., 2014), they might have less impact on the narrative skills. Narrative skill is the ability to construct a coherent and cohesive narrative text. Narrative texts are usually assessed at two levels: macrostructure, representing the global-level narrative organization, and microstructure, relating to the local sentential level. Macrostructure, reflected in the Story Grammar (sg) approach, is built upon the temporal and causal organization of events and is considered to have a more universal, i.e., language-independent nature (Iluz-Cohen & Walters, 2012; Pearson, 2002; Simon-Cereijido & Gutierrez-Clellen, 2009), whereas microstructure makes use of sentence-level linguistic units and is language-specific (Gillam et al., 2012). This chapter explores macrostructure in oral narratives of bilingual children speaking home or heritage Russian who live in Finland, Germany, Israel, Norway, and Sweden and compares it to oral narratives of monolingual children

living in St. Petersburg. Macrostructure is assessed in our study using two measures, Story Structure (SS) and Story Complexity (SC). We first compare SS and SC in the heritage Russian of children exposed to five different environmental languages (L2), Finnish, German, Hebrew, Norwegian, and Swedish, and then relate the bilinguals' narratives to narratives of monolingual children. If bilingual children exposed to different L2s perform similarly in their L1 in terms of SS and SC and bilinguals perform similarly to age-matched monolinguals, more empirical evidence of the universality of SG will be provided. Second, we scrutinize the story organization on an episodic level, specifically, we examine in detail how bilingual and monolingual children construct various episodes and combine their elements across the plot of the story. The comparisons are conducted in two age groups – younger (roughly age 4) and older (roughly age 6) participants.

Story grammar and the development of the narrative macrostructure
Macrostructure of (narrative) texts has traditionally been operationalized by means of SG (Stein & Glenn, 1979; Trabasso et al., 1989). According to this approach, narrative texts usually begin with a Setting, which introduces the characters and provides temporal and spatial background information on these characters and their environment. The Setting is followed by one or more episodes with tempo- rally and causally organized events, which are centered around a protagonist and are systematically organized. The episodes are composed of the elements, which – depending on the SG model – might vary in their type and number: e.g., Internal Event (IE) is a character's activity triggering the development of a plot; Internal response is Internal State (IS) of the character caused by the IE; Internal plans and Attempts denote the characters' strategies and actions to reach a goal; and Direct consequences and Reactions reflect characters' (non)successful results of the actions and resulting emotional states (cf. Mandler & Johnson, 1977; Peterson & McCabe, 1983, 1991; Stein & Glenn, 1979).

The mastery of SG or macrostructure implies the child's ability to use a complete set of SG elements at the episodic level to express causal relations between story events, to make inferences about characters' intentions, and to connect events to the overall thematic structure. The order of appearance of single SG elements and their combinations follows a developmental pattern. Preschool children aged 4 prefer to focus on concrete actions realized as Attempts rather than on a character's Goal, yet this might depend on the pictorial stimuli, i.e., the realization of the character's Goal depends on the extent of the Goal's accessibility during the narrative (Brown, 2007; Trabasso et al., 1992). Inclusion of Goals and IS, which express characters' intentions and the narrator's evaluation of the events, is developing later and rarely appears before the age of 8 (Khan et al., 2016; Shapiro & Hudson, 1991; Trabasso & Nickels, 1992). However, Trabasso and Nickels (1992) and recent studies using LITMUS-MAIN (special issue of *Applied Psycholinguistics*, 2016) provide evidence that children at the age of five can formulate complete GAO episodes (for similar results see Kemper, 1984; Trabasso & Rodkin, 1994).

In sum, macrostructure has mainly been assessed using the SG approach (e.g. Stein & Glenn, 1979). Even though clear developmental progression in the acquisition of macrostructure skills has been shown, research points to varying stages of acquisition for distinct SG elements and to some difficulties in the identification of single components of SG (Westby, 2012).

Story grammar models
SG has been observed to show similarities across languages (Labov, 1972). The developmental trends related to its acquisition have been reported to be similar by Berman and Slobin (1994) based on five languages and by Verhoeven and Strömqvist (2001) based on more than ten languages. Gagarina et al. (2015) reported on narratives from participants in 12 languages and 15 language pairs and found macrostructure similarity between languages in bilingual children and between bilinguals and age-matched monolinguals.

Yet some studies suggest distinctions with respect to the inclusion of the main SG elements in storytelling (Fiestas & Peña, 2004; Paul & Norbury, 2012; Warnick & Manusov, 2000). For example, the age of acquisition of an Internal Response element has been shown to differ between Cantonese- and English-speaking children as a result of culture-specific factors (Rezzonico et al., 2015). In a similar vein, Gagné and Crago (2010) raised the possibility of a potential cultural bias in the Edmonton Narrative Norms Instrument (ENNI) because SG development might not follow the same trajectory in all languages. In light of the potential effect of cross-cultural differences, the choice of narrative stimuli and coding procedure are of crucial importance in the evaluation of narrative abilities (Berman, 1995; Shapiro &

Hudson, 1991). Studies have examined narrative macrostructure skills across languages by creating methodological similarity in assessing these skills across linguistically and culturally diverse populations (e.g., Berman & Slobin's (1994) *Frog Story* project with parallel procedures for elicitation, transcriptions, coding, and analysis). However, the difficulty of assessment of macrostructure across languages and populations still remains. The challenge is in using culture-appropriate picture stimuli and a comparable theoretically based approach for the SG evaluation.

The current study follows the tradition of using a unified procedure to collect narrative data across languages. It also broadens it by using a tool with culturally adjusted stimuli created for different populations, which was piloted prior to its use. The empirical innovation of this research is in using a common set of parallel and cross-culturally robust pictorial stimuli, in which each element of macrostructure is represented, i.e. painted. The theoretical novelty of the study is in the application of the so-called multidimensional model of sg, which allows scoring of the quantitative and qualitative aspects of SG (Gagarina et al., 2012). This model differs from the SG model by Stein and Glenn (1979) in the amount, role, and specification of the episodic elements and in the evaluation of SG itself; for example, IS as the IE, and outcomes are components of the episodic structure (see Section *Method* below). The SG elements are setting, IS as IE, which introduces a problem and prompts the rest of the events in the narrative, a Goal reflecting the character's motivation to solve the problem, an Attempt to achieve the Goal, and an Outcome of the Attempt, which may or may not be successful. The IS as a reaction to the outcome is also a part of an episode (cf. Stein & Glenn, 1979; Trabasso & Nickels, 1992). Furthermore, motivated by Westby's (2005) decision tree, SG evaluation in this model is supplemented by the scoring of the episodic complexity. This scoring allows the assessment of different levels of the episodic complexity and the investigation of various sequences, which include the combination of different components, for example, the combination of Goal and Outcome denotes an incomplete episode, whereas the combination of Goal, Attempt and Outcome signifies the highest level of complexity – a complete episode. Thus, a more differentiated analysis is possible.

Assessment of macrostructure using LITMUS-MAIN

The *Language Impairment Testing in Multilingual Settings Multilingual Assessment Instrument for Narratives* (LITMUS-MAIN; Gagarina et al., 2012) was developed and piloted with over 500 children in different languages and language combinations (EU COST Action IS0804 "Language impairment in a multilingual society: Linguistic patterns and the road to assessment", Armon-Lotem et al., 2015). The LITMUS-MAIN narrative stimuli were developed to be appropriate for bilingual speakers from different cultural backgrounds (Gagarina et al., 2012). Parallel stories, each comprising three episodes, include explicit picturing of each episode's events while controlling for cognitive complexity, age of acquisition of nouns, use of mental state words, and the number of main protagonists. The depiction of events in each episode was theoretically driven in such a way that the pictorial stimuli reflect the episodes' Goals, Outcomes, Attempts, initiation, and conclusion. Thus, all components of SG are carefully presented. All studies of the current project have used similar scripts, elicitation procedures, analyses, and scoring, as laid out in LITMUS-MAIN (Gagarina et al., 2012). Analyses combine a quantitative measure of SS operationalized as a sum of all SG components and a qualitative evaluation of episodic complexity. Such an approach embraces the multidimensional nature of narrative macrostructure and allows tracing of the different ways in which children tell stories.

Recent studies have reported on monolingual and bilingual children's narratives in a variety of languages and language combinations (Bohnacker, 2016; Gagarina et al., 2015; Kunnari et al., 2016; Roch et al., 2016). Gagarina et al. (2015) showed that SG acquisition in linguistically diverse sets of monolingual (speaking 17 different languages) and bilingual children (representing 14 different language pairs, telling stories in their two languages) is similar to the previously described developmental milestones in children aged 4–10 years. Protassova et al. (2011) found that macrostructure in bilinguals ($N = 14$) is similar to that in age-matched monolinguals ($N = 22$), mean age 5;3. The special issue of *Applied Psycholinguistics* on narratives published in 2016 provides a comprehensive overview of important findings obtained with the MAIN methodology (Pesco & Kay-Raining Bird, 2016). Most studies have reported on shared macrostructure skills across bilinguals' two languages (Fichman et al., 2017; Galkina et al., 2017; Kangasaho, 2013; Rodina, 2017). SG knowledge is grounded in general cognitive development, and thus is invariant across the two languages of bilinguals tapping on the understanding of the causal and temporal relationships between protagonists of a story and their actions.

However, a certain amount of language is certainly necessary to be able to express the plot (Squires et al., 2014).

For example, fine grained analyses of SG performed by Rodina (2017) reveal cross-linguistic differences in the SS and SC. Rodina reports that overall SS scores of Norwegian-Russian bilinguals (*N* = 16, mean age 4;6) are similar in both languages. The same is found for the SC analysis, where Goals are found to be the least frequent elements in both languages while IE, Attempts, and Outcomes are the most frequent ones. AO and GAO sequences are more frequent in Norwegian. Fichman et al. (2017) examined narrative macrostructure in the narratives of Russian-Hebrew bilingual children (age 6) with typical language development (TLD) and their peers with Specific Language Impairment (SLI). The results showed that narrative macrostructure measured by SS and SC is similar across the two languages, L1/Russian and L2/Hebrew, in these two groups.

Similarly, Gagarina (2016) demonstrated strong and significant cross-linguistic correlations of SS scores in Russian-German bilinguals (preschoolers *N* = 21, mean age 3;9, first graders *N* = 15, mean age 7;0, and third graders *N* = 22, mean age 9;3). However, this finding holds only in the preschool and first-grade groups, but not in the older group of the third graders. For SC, no significant correlations between the two languages were found. Thus, the universality of SG has not received constant support, since SG is not a uniformly singular concept, but can be decomposed into different components, each having a different degree of dependence on linguistic skills, language proficiency, and various environmental factors (Gagarina et al., 2020; Kapalková et al., 2016).

In sum, MAIN allows assessment of macrostructure skills in culturally diverse populations of bilinguals. Research using MAIN stimuli and procedure supports the universality of SG but suggests an effect of language proficiency on the acquisition of specific SG elements.

Story structure and story complexity as the constituents of story grammar
Studies examining narrative skills across different bilingual age groups have shown considerable changes between the ages of 4 and 7 (Bohnacker, 2016 on English-Swedish; Gagarina, 2016 on Russian-German; Kunnari et al., 2016 on Finnish-Swedish; Roch et al., 2016 on Italian-English). The major development pertains to the complexity of an episodic unit. A well-developed episodic structure involves integrating goals as central episodic elements and thus creating complete GAO episodes. The use of goals then is a critical feature of story-telling, which takes the longest time to develop. The analytic approach utilized in MAIN allows evaluation of individual SG elements using Story Structure (SS) analysis as well as of episodic complexity implemented in Story Complexity (SC). Gagarina (2016) analyzed narrative macrostructure in the two languages of 58 Russian-German simultaneous and sequential bilinguals from 3;9 to 9;3, showing a significant improvement in both SS and SC between the ages of 3;9 and 7;0. The two older groups of children (primary school, ages 7–9) differed in their story complexity depending on the type of bilingualism: simultaneous bilinguals performed better than sequential bilinguals.

Although research shows that SG elements in each episode are organized by the character's Attempt, making Goal a central element in narrative macrostructure (Graesser et al., 1994), other constituents play an important role in narrative production as well. Gs might be not pronounced at all, but may be substituted by ISs, which trigger a protagonist's Gs, leading to actions (Gagarina et al., 2019). The identification of goals in narratives of young children can involve explicit use of motivational verbs ('want') or the use of different types of ISs and specific linguistic constructions such as verb complements. ISs assist children in expressing the causes and consequences of events and the meaning they ascribe to story characters (Norbury & Bishop, 2003). Goals and ISs reflect children's *inferencing* skills, which signal more developed story production, and are essential for a coherent narrative. Fichman et al. (2017) found that Goals were often omitted by bilingual children with typical language development and with SLI. Children from both groups used ISs instead of Goals to convey characters' motivations in L1 and L2. Thus, the use of IS contributed to the overall narrative coherence. Furthermore, research shows that different episodes may vary with respect to the inclusion of SG elements (Altman et al., 2016; Fichman et al., 2017). Fichman et al. (2017) have shown that children with language impairment include fewer elements in the first episode than children with typical language development, but they construct full second episodes.

In sum, episodic story complexity has not been sufficiency addressed in re- search on macrostructure. Different studies assign prominence to distinct SG elements. The lack of consensus may stem from variability in the stimuli and analytical approaches applied.

Thus, while the universality and developmental trajectory of narrative macro- structure have been

demonstrated by monolingual and bilingual studies arguing for cross-linguistic similarity of SG (e.g., Berman & Slobin, 1994 for monolinguals; Gagarina et al., 2015 for bilinguals) systematic analysis of macrostructure in children acquiring the same minority language spoken in different countries while applying a carefully comparable experimental design has not, to the best of our knowledge, been undertaken.

The study
The present study aims at comparing narrative macrostructure assessed in terms of two components, SS and SC, in monolingual and in L1/Russian narratives of bilingual children exposed to five different second languages – Finnish, German, Hebrew, Norwegian, and Swedish. The comparison of two age groups (4-year-olds and 6-year-olds) allows an examination of the developmental growth of narrative macrostructure. The comparison of younger bilinguals with age-matched monolinguals provides knowledge of the comparability of early monolingual and bilingual narrative skills.

The following research questions guide this study. To what extent are the two dimensions of macrostructure, SS and SC, similar across speakers of Russian with different L2s in two age groups? Does the narrative macrostructure, measured as SS and SC, of monolingual Russian speakers differ from the macrostructure of bilingual peers? How do bilingual and monolingual children in the two age groups express episodic complexity across the plot of the story – do the sequences and complexity levels differ, and if so, in what way?

Based on previous research showing similarity of macrostructure across monolinguals speaking different languages (Berman & Slobin, 1994) and across bilinguals' two languages (Fiestas & Peña, 2004; Gutiérrez-Clellen, 2002), SS and SC are predicted to be shared across speakers of Russian with different L2s within each age group, which would argue for the universality of sg. Bilinguals are not expected to differ from monolinguals (Fiestas & Peña, 2004). In line with research showing the effect of episode on macrostructure (Altman et al., 2016; Fichman et al., 2017), episodic complexity is expected to be different across episodes, thus reflecting the child's involvement in the narrative and the diversity of real-life situations, reflected in the pictorial stimuli, e.g., saving a baby from sinking/dying might have a different 'depth of threat' and impact on a story-telling as compared to bringing a baby food, because she is hungry. Therefore, these episodes might evoke various verbal descriptions.

Method

Participants
Bilingual data from 162 children living in Finland, Germany, Israel, Norway, and Sweden were elicited. Monolingual data were collected from 21 Russian children from St. Petersburg. Table 1 presents the participants' age and the bilinguals' age of onset of L2 (AoO). Bilingual participants constitute two age groups: the younger 4-year-old group ($N = 62$, mean age 54 months, age range 48–59) from Germany and Norway, and the older 6-year-old group ($N = 90$, mean age 73 months, age range 66–83) from Finland, Germany, Israel, and Sweden.

All children spoke Russian as their L1 and the language of the country as L2. All children were exposed to Russian from birth, which was spoken by at least one parent as a native language. The minimal length of exposure to L2 was at least one year (except for one German-Russian child with 9 months of exposure). 78 children were simultaneous bilinguals with AoO between 0 and 22 months and 71 children were sequential bilinguals with AoO between 24 and 60 months. The classification of these two types of bilingual was done following Ruberg (2013). Only in Germany were the bilingual participants enrolled in Russian immersion programs.

The study was performed only with those children whose parents signed pa- rental consent forms. When signing the consent form, parents were asked to complete a background questionnaire which elicited information about age, AoO, the child's socio-economic background, history of exposure to L1 and L2, patterns of language use at home, and hearing or other possible health problems, as well as parents' concerns regarding the child's language development. These questionnaires differed across the countries, but all requested information about possible deficits in hearing, motor, neurological, and cognitive development, which was collected with the aim of excluding any children with a language impairment.

Table 1. Background information

	Younger			Old	
	N	Age	AoO	Age	AoO
Finland				71.14 (66–7	0
Germany	49	54.63 (48–59)	24.21 (0–46	79.50 (76–8	15.47 (0–54
Israel				70.83 (66–7	31.45 (0–60
Norway	13	53.31 (48–58)	0		
Sweden				72.33 (67–8	11.56 (0–44
Monolinguals	21	56.67(50–59)	N/A		

AoO – Age of Onset of Bilingualism; N/A – not applicable

In order to assess the children's linguistic proficiency, the *Russian Language Proficiency Test for Multilingual Children* was performed in Germany and Israel (Gagarina et al., 2010). In these two countries, second language proficiency in German and Hebrew were assessed with two standardized tests: for German, the lexical tests WWT (Glück, 2011) and PDSS (Kauschke & Siegmüller, 2009), and for Hebrew, the Goralnik proficiency test (Goralnik, 1995). In Finland, Norway, and Sweden, the parental reports did not reveal any problems with language and motor skills for the participants in the study. In Sweden, crosslinguistic lexical tasks (CLTs; Haman et al., 2015; Ringblom et al., 2014) were also used to measure Russian and Swedish productive and receptive vocabularies. In Finland, the children's typical development was additionally confirmed by the preschool teachers. Based on the norms existing for each test in each language, bilingual children's proficiency status was confirmed. Only those bilinguals who performed above norm in at least one of their languages were included in the current study. Thus, based on the parental questionnaire, the children's performance on the proficiency tests, and conversations with the teachers (if necessary), only children with typical language development were included.

Monolingual children in St. Petersburg (there are no standardized tests for the preschool monolingual children in Russia) were tested by a commission of speech therapists, teachers, and psychologists, and their language skills were documented and described in detail. According to reports by the kindergarten teachers and specialist staff, all the children have typical motor, psycho-social, and language development.

Parents' socio-economic status ranged from low income to upper middle class, including university educators/professors. Obtaining the mean social status index for the whole cohort was not pursued due to the variety of measures across countries used to calculate this index.

Materials
The data analyzed in the present paper were elicited with the LITMUS-MAIN instrument (Gagarina et al., 2012). The LITMUS-MAIN test was developed within the framework of COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment." The stimuli include four wordless picture books and scripts of the 'Baby Birds', 'Baby Goats', 'Cat', and 'Dog' stories. All four stories have a similar structure, which is one of the main achievements of the LITMUS-MAIN design. All stories consist of six pictures each, shown to a participant in the order of a book as two, plus two, plus two pictures. The stimulus pictures and scripts begin with the Setting and are followed by three episodes. All three episodes have similar internal structure: an IS as an IE, a Goal, a character's Attempt to achieve the Goal, the Outcome of the Attempt, and an IS as a Reaction (R). The plot is similar across the stories: one of the main characters is trying to save its babies (goat) or to get food for its babies (bird), another one wants to get the babies but is chased away (cat, fox). In all countries but Israel, the 'Baby Goats' and 'Baby Birds' stories were used to elicit narrative story-telling. In Israel, the 'Dog' and 'Cat' stories were used. The structural similarity across all four stories allows comparison among them (see Figure 1 for the distribution of IS-IE, Attempt, Attempt, Attempt, IS-R).
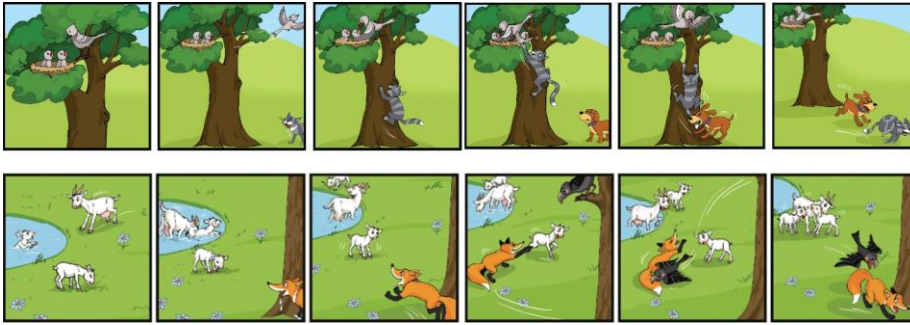
Figure 1. A comparison of the episodes of the baby birds and baby goats and the pictorial stimuli in a small-scale format (reproduced with permission from the publisher) (Gagarina et al., 2019; Gagarina et al., 2012)

Procedure

All authors employed the telling procedure suggested in the MAIN manual. Each bilingual participant narrated one of the two stories in L1/Russian (another story was narrated in L2). The stories were presented either on a computer screen or as printed books. The children were asked to choose a story in one of three envelopes and narrate it for the interlocutor in the non-joint attention mode (except for Israel, where the experimenter saw the pictures). Thus, for the majority of countries the concept of non-shared knowledge was preserved during story-telling. In Israel, shared knowledge was controlled by introducing a puppet who 'listened' to the child's narrative. The stories were similarly randomized across participants in all countries, so that half of the children did 'Cat' or 'Baby Birds' and the other half 'Dog' or 'Baby Goats'. Each child in each country was tested individually by a native speaker of Russian and each session was audiotaped. Bilinguals were tested in their respective countries of residence, while monolinguals were tested in St. Petersburg, Russia.

Transcription and coding

Prior to the analysis all narratives were transcribed verbatim by native speakers of Russian in the CLAN system (CHILDES, MacWhinney, 2000). At least 20% of the whole volume of data per country was transcribed for interrater reliability. After the transcriptions were checked, the coding was performed by the native speakers according to the protocols of the LITMUS-MAIN. If the coders of the macrostructure were not sure as to a scoring, a group of two or three researchers discussed the case until a consensus was reached.

For the goals of the present study, SG was analyzed on the basis of the multidimensional model including two main constituents, Story Structure (SS) (a quantitative measure, with a maximum score of 17 points) and Story Complexity (SC) (a qualitative measure, reflecting the well-formedness of the episode). SS was calculated as the sum of SG categories including Setting (e.g., 'once upon a time there was a nest in the tree'), IS as an IE (e.g., 'the cat saw baby birds'), Goal (e.g., 'the cat wanted to eat the baby birds'), Attempt (e.g., 'the cat climbed up the tree'), Outcome (e.g., 'the cat got one bird'), and IS as a reaction (e.g., 'the birds were scared'). Each story consisted of three episodes. Thus, there was a maximum of three ISs as an IE, three Goals, three Attempts, three Outcomes, and three ISs as a reaction. Points were assigned for each element to make up the overall score for SS. An additional two points were assigned to Setting (1 point for time and 1 point for place). Thus, the maximum SS score was 17.

The calculation of SC was operationalized in two ways, both based on the several levels of complexity, all tapping into episodic structure (cf. Westby, 1995). Three levels of complexity were identified which were related to the inclusion of SG elements in an episode. The first and highest level of episodic complexity was achieved when a child produced the sequence of all three main SG elements, that is, Goal-Attempt-Outcome (GAO). In this first analysis the child receives 1 point for the production of the complete GAO episode, whereas failure to produce the complete GAO receives 0. A second level of SC coding targeted incomplete episodic structure in which a child included Goal. A child received a score of 1 if she produced Goal-Attempt (GA) or Goal-Outcome (GO). Since young children (and adults) tend to omit goals (Trabasso et al., 1992), the SC analysis also included scoring of incomplete episodic structures where Goal was omitted, Attempt-Outcome (AO). Absence of any of these sequences would receive 0. The numbers of episodes containing complete episodes GAO, incomplete episodes GA/GO, and AO sequences were calculated. The choice of this specific coding system for SC allowed us to obtain the full repertoire of children's performance at the episodic level. The three approaches to coding SC were

inclusionary; that is, a child mentioning a Goal, an Attempt, and an Outcome would receive 1 for GAO, 1 for the GA/GO level, and 1 for the AO level. In other words, children received scores for all three levels of SC. The rationale behind using this scoring system was to understand which level of SC is best reflected in children's narratives. While GAO represents the completeness of the episode, GA/GO reflects the inclusion of the Goal, and AO reflects episodes where a child produces necessary events but omits the motivation behind the characters' actions.

Analysis
Although the present research does not provide data for the two age groups across all countries, narrative data for the younger and the older groups were analyzed separately due to the crucial developmental progression/trend in macrostructure from the age of four to the age of seven. Furthermore, 4-year-old monolinguals were used as a control group in order to provide evidence for the development of macrostructure skills. If bilinguals receive lower scores than monolinguals, we wanted to know whether the gap remains in the older bilingual group.

Results
Macrostructure was analyzed using two components, SS, which reflected the sum of the main SG elements, and SC, which focused on the episodic complexity. The analyses aimed to examine narrative macrostructure in two age groups of speakers of Russian as a first language living in different countries and to compare bilinguals' performance to that of monolinguals.

Story structure
The analysis first compared the SS score for the two age groups. For children across all countries, a significant difference in SS score between the younger and the older age groups was revealed using an independent t-test, $t(150) = -7.92$, $p < .001$. For the younger group, the analyses examined children from Germany and Norway, and for the older age group, results were obtained for the children from Germany, Finland, Israel, and Sweden. Based on these findings, all further analyses were conducted separately for the two age groups. Table 2 presents the mean SS score obtained from bilingual participants per country for the two age groups. Monolingual children from Russia ($N = 21$) had a mean SS score of 7.67 ($SD = 2.56$). In order to test the effect of the type of story, narratives retold to the Cat/Dog stimuli were compared to the narratives elicited using the Baby Birds/Baby Goats stimuli. For SS, an independent sample t-test revealed no significant difference between the two types of stories, $t(171) = 1.72$, $p = .09$. Since bilingual children varied in AoO (see Table 1), an analysis was performed comparing simultaneous and sequential bilinguals, which revealed no significant differences between the groups on their performance on SS, $t(147) = -.09$, $p = .93$.

Table 2. Number of children, mean SS score, and standard deviation per age group

	Younger			Older	
	Mean	SD		Mean	SD
Finland				6.29	1.70
Germany	4.29	2.01		8.39	1.85
Israel Norway			47	6.72	2.02
Sweden	5.38	1.5			
			18	6.83	1.15

To investigate to what extent children in different countries used SG categories, the SS score was submitted to two one-way ANOVAs for the younger and the older age groups. For the younger group, the analysis did not reveal a significant group difference, $F(1, 61) = 3.37$, $p = .07$. Children in Germany and in Norway included a similar number of SG elements. For the older group, the analysis revealed a significant effect, $F(3, 86) = 4.24$, $p < .01$, $\eta^2 = .13$, which was due to the higher SS score received by Russian speakers from Germany, as revealed by post-hoc Bonferroni analyses. Bilinguals in Germany used more SG elements than bilinguals in Israel ($p = .009$), but a difference between Germany and Finland ($p = .07$), Germany and Sweden ($p = .07$), Finland and Israel ($p = 1.0$), and Finland and Sweden ($p = 1.0$) was not observed.

Monolingual Russian-speaking children were first compared to age-matched young bilingual groups. A one-way ANOVA revealed a significant group difference, $F(2, 82) = 19.12$, $p < .001$, $\eta^2 = .2$, where monolinguals obtained a higher SS score than bilinguals from Germany ($p < .001$) and from Norway ($p < .01$) as revealed by Bonferroni post-hoc tests. In order to see whether the monolingual-bilingual difference extends to older children, the SS scores of monolinguals and older bilinguals were submitted to analysis. A one-way ANOVA did not reveal a significant group difference between monolinguals and older bilingual children from all countries. To summarize, younger bilinguals performed similarly on SS across countries, whereas in the older bilingual group only speakers from Germany significantly outperformed bilinguals from Israel. Monolinguals outperformed age-matched bilinguals but performed similarly to older bilinguals. To further investigate narrative macrostructure on the episodic level and to explore the sources of group
differences, we analyzed SC.

Story complexity
SC was analyzed in two ways. First, we examined the number of complete episodes including a Goal, an Attempt, and an Outcome (GAO), where each child could have between 0 and 3 GAOs. Second, we registered the number of episodes containing sequences, such as GA/GO or goalless episodes, including just AO. This coding was motivated by previous research indicating that children and adults do not always use all three elements in every episode; in particular, G might not be always present, and sequences of IS as IE together with AO may still portray the full episode (Gagarina et al., 2019 on adult MAIN narratives). Examining SG sequences in addition to complete GAO episodes provides a more differentiated view of the level of SG and an insight into children's macrostructure skills. Table 3 displays the mean number of episodes containing a complete GAO, the number of episodes containing a GA or GO sequence, and the number of episodes containing an AO sequence. Since the data on SC were measured on an interval scale, they were analyzed for across-group comparisons (across countries for each age group and monolinguals vs. bilinguals) using non-parametric statistical tests.

Table 3. Mean number (and standard deviation) of episodes containing GAO, GA/GO, and AO

	Younger			Older		
	GAO	**GA/GO**	**AO**	**GAO**	**GA/GO**	**AO**
Finland				0.29 (0.49)	0.86 (0.90)	0.71 (0.49)
Germany	0.14 (0.35)	0.49 (0.65)	0.27 (0.49)	0.50 (0.62)	0.22 (0.43)	0.39 (0.70)
Israel				0.51 (0.72)	0.55 (0.83)	0.43 (0.58)
Norway	0.15 (0.38)	0.38 (0.65)	0.31 (0.48)			
Sweden				0.28 (0.46)	0.61 (0.70)	0.83 (0.62)

For the effect of story type (Cat/Dog vs. Baby Birds/Baby Goats), a Mann-Whitney test revealed no significant difference for GAO, $U = 2705.00$, $p = .13$, GA/ GO, $U = 2936.50$, $p = .59$, or AO, $U = 3059.00$, $p = .95$. For the difference between simultaneous and sequential bilinguals, a Mann-Whitney test did not reveal significant results, $U = 2722.50$, $p = .82$.

In the younger group, a Mann-Whitney test did not reveal a significant difference between children from Germany and Norway for GAO, $U = 315.00$, $p = .92$, and for the GA/GO sequence, $U = 288.50$, $p = .55$, or the AO sequence, $U = 300.50$, $p = .68$. For the older group, a Kruskal-Wallis H test for multiple groups showed that children from Germany, Finland, Israel, and Sweden used a similar number of GAO episodes, $\chi^2(3) = 1.72$, $p = .63$, and for the GA/GO sequence, $\chi^2(3) = 4.66$, $p = .20$. A significant group difference emerged for the AO sequence, $\chi^2(3) = 8.71$, $p < .05$. Post-hoc tests aiming to reveal the source of the group effect for AO failed to show significant differences between countries (the α level was set to .008 using Bonferroni corrections).

Monolinguals had a higher number of episodes including GAO ($M = 0.67$), the GA/GO sequence ($M = 0.60$), and the AO sequence ($M = 0.66$) compared to bilinguals. Two Kruskal-Wallis tests were performed comparing monolinguals first to the younger and then to the older bilinguals. For the younger bilingual group, a significant effect emerged for GAO, $\chi^2(2) = 15.77$, $p < .001$, and the AO sequence, $\chi^2(2) = 7.50$, $p < .05$, but no difference emerged for the GA/GO sequence, $\chi^2(2) = 0.40$, $p = .82$. In order to test the source of the effect, we performed a series of Mann-Whitney tests setting the α level to .017

for the group. For GAO, Mann-Whitney tests confirmed that monolinguals used more complete GAOs than children from Germany, $U = 287.00$, $p < .001$, but the difference was not confirmed for children from Norway, $U = 77.50$, $p = .035$. For AO, monolinguals used more sequences than children from Germany, $U = 340.50$, $p = .007$, but the difference was not significant for children from Norway, $U = 96.50$, $p = .16$.

For the comparison between monolinguals and the older bilingual group, Kruskal-Wallis tests did not reveal any difference between monolinguals and bilinguals for GAO, $\chi^2(3) = 4.28$, $p = .37$, and GA/GO, $\chi^2(3) = 4.88$, $p = .30$. For AO, the difference was marginally significant, $\chi^2(3) = 9.47$, $p = .05$. Post-hoc tests did not reveal significant differences between monolinguals and older bilinguals from Finland, Germany, Israel, and Sweden.

To explore the role of Goals, we examined which of the three episodes contained explicit naming of all three elements, Goal, Attempt, and Outcome, and which of them included IS as IE, instead of Attempt, in combination with Attempt and Outcome. This type of analysis aimed at providing insights into the episodic structure and how the Goal information is stored in each episode. The data from different countries were collapsed to perform this analysis. Data from Israel was excluded for the comparison across episodes because a different set of pictures was used with Russian-Hebrew bilinguals (Dog and Cat stories). The comparison was made between younger bilinguals, age-matched monolinguals, and older bilinguals. Initially, we compared the three episodes for the inclusion of Goal information. All children included fewer Goal sequences in the first episode than in the second or the third episodes. When the data were collapsed across the age groups and countries, a significant cross-episode effect emerged, $\chi^2(2) = 34.84$, $p < .001$. A Wilcoxon test revealed that there were significantly fewer Goal-related sequences in the first episode than in the second episode, $Z = -4.42$, $p < .001$, or the third, $Z = -5.00$, $p < .001$. The difference between episodes two and three was not significant, $Z = -.22$, $p = .83$. These results illustrate that different episodes trigger a different quantity (and possibly quality) of SG elements. To explore this further, we compared complete GAO and AO sequences in each episode. We suggested that IS, encoded as IE, might implicitly provide goal information and enrich episodes when the explicit goal is omitted, thus generating an episode with ISAO structure. For example, instead of saying 'Mother bird wants to get the food for her chicks, flies away and brings a worm' – which is the full GAO episode, a story-teller might say 'The chicks are hungry, so their mother bird flies away and brings a worm' – which is an ISAO episode and is not less 'valuable' in terms of complexity. Actually, the word 'hungry' implies, given the world knowledge, that if a baby is hungry, the mother's task is to get the food.

Figure 2 shows the use of complete GAO and ISAO in each of the three episodes by younger bilinguals, older bilinguals, and monolinguals. As seen in Figure 2, the use of complete GAO was rare and restricted mainly to the second episode of monolinguals. The use of a complete GAO sequence is illustrated in (1), where all three episodic elements, Goal, Attempt, and Outcome, are mentioned (all examples are glossed, and the participants' age are provided in the brackets).

(1) 'Baby birds', episode 2, GAO:
Attempt 2. kot lez na derevo.
      cat-MASC.NOM climbed-MASC.IMPF on tree-NEUT.ACC
      'The cat was climbing the tree.'
Goal 2.   i xotel est' ptenčikov.
      and wanted-MASC.IMPF eat-INF baby-birds-ACC
      'And (he) wanted to eat the baby-birds.'
      mama ne zametila
      mother-FEM.NOM NEG noticed-FEM.PF
      'Mother did not notice'
Outcome 2. čto xitryj kot čut' ne
      that mean-MASC.NOM cat-MASC.NOM almost NEG
      lovil ptenčika.
      catch-MASC.IMPF baby-bird-MASC.GEN
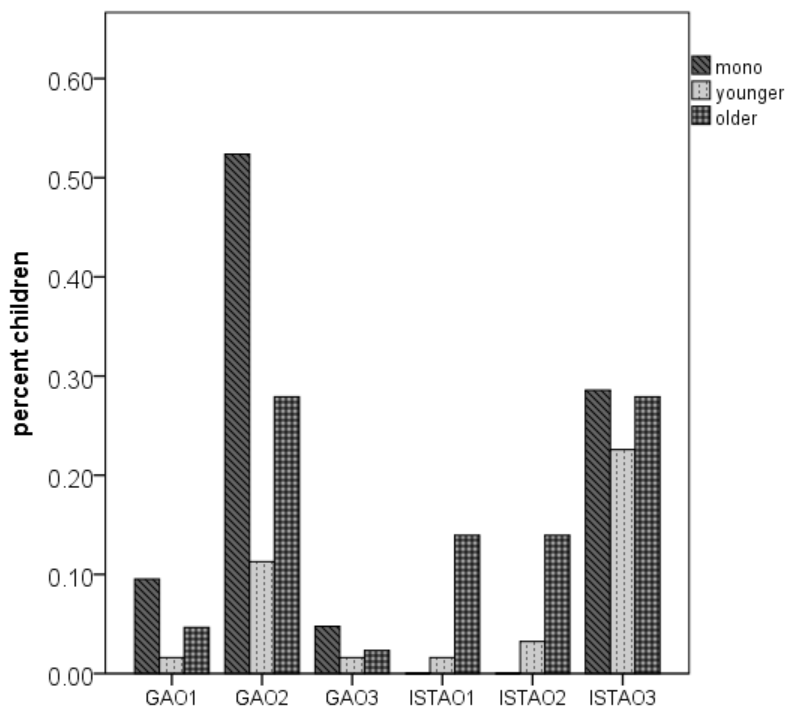   'that the mean cat almost ate the baby-bird.'   (6;1)

Figure 2. Complete GAO episode and ISAO per episode

Friedman non-parametric tests comparing the use of GAOs across the three episodes in younger bilinguals, older bilinguals, and monolinguals yielded a significant difference for younger bilinguals, $\chi^2(2) = 8.00$, $p = .02$, older bilinguals, $\chi^2(2) = 15.86$, $p < .001$, and monolinguals, $\chi^2(2) = 15.17$, $p < .001$. Further pairwise comparisons were conducted using Wilcoxon tests; the α level was set to 0.02. The analyses revealed that more GAOs were included in episode 2 than in episode 1 in the older bilingual group, $Z = -2.89$, $p = .004$, and in monolinguals, $Z = -2.71$, $p = .007$, and more GAOs were included in episode 2 than in episode 3 in the older bilingual group, $Z = -3.05$, $p = .002$, and in monolinguals, $Z = -3.16$, $p = .002$. The differences across the episodes did not reach significance in the younger bilingual group.

Example (2) illustrates the use of SG sequences including an IS, Attempt, and Outcome, which provide all the necessary information to expresses the pictorial stimuli and can be analyzed as a complete episode. The IS describing that the cat saw the baby birds provides implicit information about the cat's intentions. Such sequences were more consistent across episodes than GAO. Children in the younger group had a significant across-episode difference, $\chi^2(2) = 24.50$, $p < .001$, such that episode 3 triggered more ISAO sequences than episode 1, $Z = -3.74$, $p < .001$, or episode 2, $Z = -3.50$, $p < .001$ (see Figure 2), while children in the older group and monolingual children produced a similar number of ISAO sequences in all three episodes.

(2)'Baby birds', episode 2, ISAO

IS 2.   deti   vidjat   košku.
children-NoM see-3pL cat-FeM.Acc
'The children see the cat.'
koška   vidit   ix.
cat-FeM.NoM see-3 SG them
'The cat sees them.'
a   mama   uletela.
and mother-FeM.NoM flew-FeM.pF
'And mother flew away.'

Attempt 2.   koška   načala   zalezat'.
cat-FeM.NoM began-FeM.pF climb-INF.IMpF
'The cat started climbing.'
a   mama   davala   červjaki.
and mother-FeM.NoM gave-FeM.IMpF worms-NoM
'And mother was giving worms.'

Outcome 2.   koška   dostala   odnogo
cat-FeM.NoM reached-FeM.pF one-MAsc.Acc

ptenčika.
baby-bird-MAsc.Acc
'The cat got one baby-bird.' (5;7)

In sum, a complex interaction of SG sequences and full or complete episodes emerged in the three groups. First, complete GAO episodes collapsed across participants ($M = 0.36$) were less frequent compared to ISAO sequences ($M = 0.53$). GAO sequences appear mainly in the second episode and more so in monolinguals, and ISAO sequences appear frequently in the second and the third episodes. The third episode triggered the most ISAO sequences in the three groups. Thus, except for the complete GAO episode, the two sequences show an increase in frequency towards the end of the narrative. Crucially, older bilinguals demonstrate less variation among the three episodes, which might indicate a developmental trend towards stability in incorporating all parts of a narrative.

Discussion

The current study investigated macrostructure in elicited narratives of bilingual children speaking heritage Russian as their first language and compared them to narratives produced by monolinguals. Bilingual narrative data were collected in Finland, Germany, Israel, Norway, and Sweden and monolingual data were collected in St. Peterburg by means of the LITMUS-MAIN using similar sets of pictorial stimuli, elicitation procedures, and scoring. Bilingual children were investigated in two age-groups, roughly four and six years old, and monolingual children were age-matched to the younger group. Macrostructure was assessed using a multidimensional system, combining two scoring levels. SS provides quantitative information about the overall inclusion of the main SG elements. SC reflects a qualitative measure by means of encoding the three main SG components GAO or their equivalent, i.e., ISAO, which allows the assessment of the episode-level techniques applied by children to convey narrative coherence. The research pursued three main goals: (1) it compared narrative skills in heritage speakers of Russian as L1 who are exposed to different second languages and thus explored the universality of narrative skills; (2) it compared bilingual children's macrostructure performance to that of monolingual children; and (3) it scrutinized the episodic structure of the narratives' plot revealed by the two SC measures, GAO and ISAO in two age groups (4-year-olds and 6-year-olds).

First, the comparison of narrative story-telling skills in Russian-speaking bilinguals living in Germany, Norway, Finland, Israel, and Sweden revealed a number of cross-country similarities which highlight the robust nature of the children's narrative abilities across those countries and shows evidence for the universality of sg. We found that within each age-matched group, bilinguals show similar SS and SC scores, except for the difference between children from Germany and Israel on SS. The low SS scores in the younger age group (speakers of heritage Russian from Germany and Norway) indicate that they included fewer SG elements (4.29 for Germany and 5.38 for Norway out of 17) than the older age group (speakers of Russian from Germany, Finland, Israel, and Sweden). These results corroborate previously made observations that story-telling abilities of four-year-olds are in the initial stages of development (e.g., Berman & Slobin, 1994; Peterson & McCabe, 1983) and the diversity in the younger age, in our case the onset of story-telling acquisition at age four, might be stronger two or three years later when children's narrative skills are already advanced (see Bohnacker & Gagarina, 2020 on story comprehension).

The findings confirm the results of previous studies showing that SG is similar in speakers of different languages provided that they share similar cultural conventions about language (Berman & Slobin, 1994; Verhoeven & Strömqvist, 2001). Berman and Slobin (1994) showed how the choice of linguistic forms is closely related to the functions they serve in narrative. Similar development of form–function relations was observed across nine different languages. Developmentally, similar milestones in narrative production were observed across languages in children ages 3–9 telling the *Frog Story* and also in previous studies using MAIN for story elicitation and comprehension (see the special issues on MAIN such as *Applied Psycholinguistics*, 2016; *First Language*, in press and *Linguistic Approaches to Bilingualism*, in press).

A considerable difference in the inclusion of SS elements in narratives was observed between the age-matched children in Germany and Israel, but children in all countries performed similarly on SC. In Germany, an extensive network of children's bilingual educational institutions with an intensive support of the home language Russian (as well as governmental integration policies, see Walters et al., 2014) and educational programs including book-reading and story-telling might lead to generally higher linguistic

skills, which is reflected in the number of SG components produced in the narratives (cf. Kapalková et al. (2016) indicating that language proficiency plays an important role in story structure production). In order to produce a higher number of the story structure elements one might need more language. The low performance on SS of the children from Israel could be explained by their dominance in L2/Hebrew. The results of the present research show that quality of exposure to story-telling in combination with explicit literacy instruction, which children were exposed to in Germany, may enhance the mac- rostructure quantity, i.e., the number of SS components which a child produces. Conversely, insufficient exposure to literacy instruction might weaken (quantitative) narrative performance.

The lack of difference between the participant groups across countries in SC indicates that the two elements, SS and SC, reveal different developmental profiles in narrative performance. SS relates to the number of included SG elements and is associated more strongly with linguistic skills, as a more quantitative measure, whereas SC reflects the complexity of the smallest composite units of a plot, i.e., episodes. Children might include fewer elements in the production of SS due, for example, to insufficient vocabulary, but the complexity of the episodic structure seems to be affected by proficiency to a lesser extent. Basically, if a child understands the causal relationship of an episode, i.e., the intention/goal to do something – action – result, s/he will be able to express this triple with fewer linguistic means or perhaps with less exact wording, while still revealing the essence of it.

Research has suggested that the number of SG elements does not fully reflect macrostructure skills, and attempts have been made to use other episode-oriented approaches, such as the hierarchical goal plan of the character(s) (Trabasso et al., 1992).

Bilinguals vs. monolinguals
The comparison of bilinguals and monolinguals revealed significant differences for the SS scores in four-year-olds. Monolinguals outperformed the age-matched bilinguals and demonstrated scores that were rather similar to the older bilingual age group. In our view, the observed differences do not necessarily indicate bilingual disadvantage, since Russian-speaking bilinguals and monolinguals may have very different educational, social, and cultural experience with story-telling. The experience of Russian monolinguals may be considerably richer due to the intensive educational programs in the kindergartens in St. Petersburg, which involve explicitly teaching story-telling skills. As previous studies have shown, teaching narrative skills results in considerable improvement in this ability in preschool children (Hayward & Schneider, 2000). Similarly, Gagarina (2016) has shown that when the bilingual school program includes explicit teaching of story-telling (structure) in the home language (in this case Russian), SC in this language is higher compared with the school language (in this case German). In her study, the third grade primary school children were significantly better in SC in Russian as com- pared to German, whereas no such relation was found for SS. Therefore, it would be more appropriate to draw parallels between bilinguals and monolinguals in the same country of residence and account for the content of language development programs for children.

Episodic structure

Detailed analysis of the highest attainment of SC, as represented by GAO and ISAO (see Figure 2), revealed that 6-year-old bilinguals and 4-year-old monolinguals performed significantly better than 4-year-old bilinguals in all episodes. This suggests that bilinguals' narrative production develops prior to age 6 and that the developmental path may be more protracted than that of monolinguals. A comparison of the denotation of the specific episodes revealed that older bilinguals and monolinguals produced significantly more GAOs in the second episode, which might be said to be the culmination or the core of the story and the most emotional part – it is the intent of the cat/fox to eat up a baby. The final episode is the rescue of the baby from being eaten.

While Goals are highly infrequent in children's productions across all groups, we argue that the well-formedness of the episodes is nevertheless sustained through the use of the ISs in combination with AO sequences. In the absence of the explicit goal elements, ISs are frequently employed by four- to six-year-old children to encode the characters' intentions and to state problem.

Additionally, we suggest that while evaluating the complexity of macrostructure in the Russian-speaking children's narratives, one should consider language-specific peculiarities and verb telicity. More specifically, the rich inflectional paradigm of verbs in Russian may provide many opportunities for

children to express a variety of actions encoded as attempts and outcomes. Namely, Russian marks aspect on most verbs, which has a particular relevance for narrative context, where completeness of action and telicity are reflected in the choice of verb. For example, in *sobaka uvidela i stjanula košku za xvost* (dog saw-FeM.pF and pulled.down-FeM. pF cat-Acc by tail-Acc), both telic verbs exhibit perfective aspect and specific Aktionsart, which have inherent completeness. Thus, similar to the coding system proposed by Trabasso and colleagues, where goal information can be inferred from actions or outcomes (Trabasso & van den Broek, 1985; Trabasso et al., 1989), the Attempt-Outcome information is encoded within a single telic perfective verb form *stjanut'* 'pull-down-PERF.PAST', which denotes an endpoint of a durative telic action.

Finally, the in-depth analysis of the narratives also revealed that the three episodes differ in the number of SG components they trigger. Namely, younger vs. older bilinguals used different techniques to verbalize the episodes' content as far as the production of the complete GAO and ISAO was concerned (see Figure 2). We suggest that this finding may have important theoretical and practical implications. In particular, episodes 2 and 3 appear to be more elaborated and more often include the complete GAO or ISAO than episode 1 across all countries. This is in line with previous observations, which offer two possible explanations for why episode 1 (helping a baby goat or feeding baby birds) receives less attention (cf. Altman et al., 2016; Fichman et al., 2017). First, the processing load may be higher at the start when a child is expected to realize and formulate an IE which is supposed to trigger all subsequent events in the story. This finding can be implemented in clinical and educational settings, where children should receive more detailed instructions on how to initiate narration, which might also be crucial for relating all subsequent episodes. Second, the emotional involvement of children in the central (second) episode may be higher, since it depicts how the mean cat climbs and grabs the baby birds and how the fox catches a baby goat. The higher emotional involvement of the narrator may thus result in a more detailed elaboration of the events in episode 2. To conclude, the difference in the emotional load, the intensity, and the importance of actions triggered by different episodes should be taken into account when per- forming the macrostructure analysis of multi-episodic narratives.

Yet a different picture emerged when GAO was compared with ISAO per episode. While GAO was most frequently used by younger bilingual and monolingual children in the second episode, ISAO was most numerous in these children in the third episode. In particular, about one-third of the older bilingual children produced the Goal (in GAO) for episode 2: Cat or Fox (episode characters: cat and birds or fox and baby goat) "Cat/Fox wanted to eat/catch/kill the/a baby bird(s)/ baby goat" (Gagarina et al., 2012, pp. 123, 129), but IS as an IE (in ISAO) for episode 3: Dog or Bird (episode characters: dog, cat, and birds or bird, fox, and baby goat) "Bird saw that the goat was in danger or Baby goat was in danger; Dog saw that the bird was in danger / that the cat caught/got the bird" (Gagarina et al., 2012, pp. 124, 130). This might be explained by the nature of the depicted events: while in *eating* and *killing* of another animal children verbalize this very action and not the environmental factors, the mother-bird/-goat leaves the babies alone and doesn't notice the danger (episode 2), in the situation of *saving of a victim* the fact of (the observation of) danger is verbalized. These results indicate that ISAO, which implicitly encodes Goal information without an explicit goal statement, might reflect the developmental stage in the acquisition of sg. Namely, children might reach the higher level of SC via ISAO while omitting Goals and the full episode doesn't necessarily include the Goal.

Conclusion

The main conclusion of the present study is related to the similarity of macrostructure investigated in two age groups of preschool bilinguals with L1 Russian and five different L2s and operationalized by two main measures, SS and SC. First, we have been able to show that narrative macrostructure in L1/Russian is similar across age-matched bilingual groups exposed to different L2s, with the one exception of Russian-German bilinguals in the older age group outperforming age-matched children in Israel on the SS measure, despite the uneven distribution of participants across countries and the lack of balance across age groups. Secondly, the current results indicate the importance of analyzing SC in narrative evaluation and suggest that besides the widely accepted complete GAO unit, the mastery of episodic structure involves the combination of the AO sequence with IS, which contributes to a coherent story. This is because this latter combination, equally to GAO, indicates the protagonist's intentions and delineates the most complex and complete episodes; thus, it can be seen as an indicator of fully acquired narrative skills. Finally, we found that children behave differently in narrating different episodes. This latter finding has

especially important implications for the evaluation of narrative skills in children, since it presents evidence for the importance of pictorial stimuli and calls for caution in the choice of pictorial stimuli when assessing (bilingual) children's narratives skills. All in all, this study contributes to our knowledge about the acquisition of narrative skills in bilingual children, enriches our understanding about L1/Russian, and provides insights into the variety of verbalization strategies applied by children across episodes in wordless picture stories.

References
Aleshkovski, I., Grebenyuk, A., & Vorobyeva, O. (2018). The evolution of Russian emigration in the post-Soviet period. *Social Evolution & History*, 17(2), 140–155. https://doi.org/10.30884/seh/2018.02.09

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, 37(1), 165–193. https://doi.org/10.1017/S0142716415000466

Amit, K. (2010). Determinants of life satisfaction among immigrants from Western countries and from the FSU in Israel. *Social Indicators Research*, 96(3), 515–534. https://doi.org/10.1007/s11205-009-9490-1

Arefiev, A. L. (2017). *Actual state and the tendencies of spread of the Russian in the world. Scholarly publications*, G. V. Osipov (Ed.). The Institute of Socio-Political Research RAS.

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Multilingual Matters. https://doi.org/10.21832/9781783093137

Armon-Lotem, S., Walters, J., & Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguistic Approaches to Bilingualism*, 1(3), 291–317. https://doi.org/10.1075/lab.1.3.04arm

Berman, R. A. (1995). Narrative competence and storytelling performance: How children tell stories in different contexts. *Journal of Narrative and Life History*, 5(4), 285–313. https://doi.org/10.1075/jnlh.5.4.01nar

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Lawrence Erlbaum Associates.

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U., & Gagarina, N. (Eds.). (2020). *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives* (Studies in Bilingualism 61). John Benjamins. https://doi.org/10.1075/sibil.61

Brown, D. (2007). Young children's use of causal connections during storytelling: The role of context and individual differences in attention (Unpublished PhD dissertation). University of Louisville.

Fichman, S., Altman, C., Voloskovich, A., Armon-Lotem, S., & Walters, J. (2017). Story grammar elements and causal relations in the narratives of Russian-Hebrew bilingual children with SLI and typical language development. *Journal of Communication Disorders* 69, 72–93. https://doi.org/10.1016/j.jcomdis.2017.08.001

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, 35(2), 155–168. https://doi.org/10.1044/0161-1461(2004/016)

Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: Rasskaz and Erzaehlung. *Applied Psycholinguistics*, 37(1), 91–122. https://doi.org/10.1017/S0142716415000430

Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., Golcher, F., & Walters, J. (2014). Age, input quantity and their effect on linguistic performance in the home and societal language among Russian-German and Russian-Hebrew preschool children. In R. K. Silbereisen, P. F. Titzmann, & Y. Shavit (Eds.), *The challenges of diaspora migration. Interdisciplinary perspectives on Israel and Germany* (pp. 63–82). Ashgate.

Gagarina, N., Bohnacker, U., & Lindgren, J. (2019). Macrostructural organization of adults' oral narrative texts. *ZAS Papers in Linguistics*, 62, 190–208. https://doi.org/10.21248/zaspil.62.2019.449

Gagarina, N., Klassert, A., & Topaj, N. (2010). Sprachstandstest Russisch für mehrsprachige Kinder/Russian language proficiency test for multilingual children. *ZAS Papers in Linguistics*, 54. https://doi.org/10.21248/zaspil.54.2010.402

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 bilingual- ism from language impairment* (pp. 243–276). Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gagarina, N., Lomako, J., Stadtmiller, E., & Lindner, K. (2020). Text organization in typically developing bilinguals and bilinguals at risk of DLD: what is different and how language independent is it? In N. Gagarina & R. Musan (Eds.), *Referential and relational discourse coherence in adults and children* (pp. 85–104). Berlin: De Gruyter. https://doi.org/10.1515/9781501510151-005

Gagné, A., & Crago, M. (2010). The use of the ENNI to assess story grammar competency of school-aged French speaking children with and without specific language impairment. *Canadian Journal of Speech-Language Pathology and Audiology*, 34(4), 231–245.

Galkina, E., Ringblom, N., Krasnoshchekova, S., & Rogozhkina, G. (2017). *Acquiring gender and case in Russian: Russian-Dutch, Russian-Swedish and Russian-Azerbaijani bilingual children compared with their monolingual peers with and without SLI and adult learners of Russian as a foreign language*. Poster presented at 14th International Congress for the Study of Child Language, Lyon, France.

Gillam, S. L., Gillam, R. B., & Reece, K. (2012). Language outcomes of contextualized and decontextualized language intervention: Results of an early efficacy study. *Language, Speech, and Hearing Services in Schools*, 43(3), 276–291. https://doi.org/10.1044/0161-1461(2011/11-0022)

Glück, C. W. (2011). *WWT 6–10, Wortschatz- und Wortfindungstest für 6- bis 10-Jährige* (2nd ed.). Elsevier.

Goralnik, E. (1995). *Goralnik Diagnostic Test*. Matan.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13(2), 175–197. https://doi.org/10.1016/S0898-5898(01)00061-4

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.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Multilingual Matters. https://doi.org/10.21832/9781783093137-010

Hayward, D., & Schneider, P. (2000). Effectiveness of teaching story grammar knowledge to pre- school children with language impairment: An exploratory study. *Child Language Teaching and Therapy*, 16(3), 255–284. https://doi.org/10.1177/026565900001600303

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual pre- school children with typical and impaired language. *Bilingualism: Language and Cognition*, 15(1), 58–74. https://doi.org/10.1017/S1366728911000538

Kangasaho, E. (2013). Razvitie narrativnyh sposobnostej u dvujazyčnyh detej doškol'nogo vozrasta [The development of narrative abilities in preschool bilinguals] (Unpublished MA thesis). University of Helsinki.

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454

Kauschke, C., & Siegmüller, J. (2009). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)* (2nd ed.). Elsevier.

Kemper, S. (1984). The development of narrative skills: Explanations and entertainments. In S. A. Kuczaj (Ed.), *Discourse development: Progress in cognitive development research* (pp. 99–124). Springer. https://doi.org/10.1007/978-1-4613-9508-9_5

Khan, K. S., Gugiu, M. R., Justice, L. M., Bowles, R. P., Skibbe, L. E., & Piasta, S. B. (2016). Age-related progressions in story structure in young children's narratives. *Journal of Speech, Language, and Hearing Research* 59(6), 1395–1408. https://doi.org/10.1044/2016_JSLHR-L-15-0275

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. https://doi.org/10.1017/S0142716415000442

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates. https://doi.org/10.1177/026565909200800211

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9(1), 111–151. https://doi.org/10.1016/0010-0285(77)90006-8

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, 38(3), 287–313. https://doi.org/10.1080/136820310000108133

Parkvall, M. (2016). *Sveriges språk i siffror: Vilka språk talas och av hur många?* Morfem & Språkrådet.

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence* (4th ed.). Elsevier.

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). Multilingual Matters. https://doi.org/10.21832/9781853595721-008

Pesco, D., & Kay-Raining Bird, E. (2016). Perspectives on bilingual children's narratives elicited with the Multilingual Assessment Instrument for Narratives. *Applied Psycholinguistics*, 37(1), 1–9. https://doi.org/10.1017/S0142716415000387

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. Plenum. https://doi.org/10.1007/978-1-4757-0608-6

Peterson, C., & McCabe, A. (1991). Linking children's connective use and narrative macrostructure. In A. McCabe & C. Peterson (Eds.), *Developing narrative structure* (pp. 29–53). Lawrence Erlbaum Associates.

Protassova, E., Petrovskaja, V., & Ovchinnikov, A. (2011). Opisanie kartinok finsko-russkimi bilingvami. In S. N. Ceytlin & M. B. Eliseeva (Eds.), *Put' v jazyk: odnojazychie i dvujazychie* (pp. 282–293). Jazyki slavjanskix kul'tur.

Remennick, L. (2003). A case study in transnationalism: Russian Jewish immigrants in Israel of the 1990s. In R. Münz & R. Ohliger (Eds.), *Diasporas and ethnic migrants: Germany, Israel, and post-Soviet successor states in comparative perspective* (pp. 370–384). Frank Cass.

Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., Milburn, T., Pelletier, J., Weitzman, E., & Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language & Communication Dis- orders*, 50(6), 830–841. https://doi.org/10.1111/1460-6984.12179

Ringblom, N., Håkansson, G., & Lindgren, J. (2014). *Cross-linguistic lexical tasks: 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(1), 49–67. https://doi.org/10.1017/S0142716415000417

Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *Inter- national Journal of Bilingualism*, 21(5), 617–635. https://doi.org/10.1177/1367006916643528

Ruberg, T. (2013). Problembereiche im kindlichen Zweitspracherwerb. *Sprache–Stimme–Gehör*, 37(4), 181–185. https://doi.org/10.1055/s-0033-1358698

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, 27(6), 960–974. https://doi.org/10.1037/0012-1649.27.6.960

Simon-Cereijido, G., & Gutierrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evalua- tion of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, 30(2), 315–337. https://doi.org/10.1017/S0142716409090134

Squires, K. E., Lugo-Neris, M. J., Pena, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language and Communication Disorders* 49(1), 60–74. https://doi.org/10.1111/1460-6984.12044

Statistics Finland (2017). *Population 31.12.2016 by Region, Language, Age, Sex and Year*. Statis- tics Finland's PX-Web databases. Retrieved from https://pxnet2.stat.fi/PXWeb/pxweb/en/ StatFin

Statistics Norway (2017). Available online at: http://www.ssb.no

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Ablex.

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. https://doi.org/10.1080/01638539209544812

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 cross-linguistic developmental study* (pp. 85–106). Lawrence Erlbaum Associates.

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. https://doi.org/10.1016/0885-2014(92)90009-G

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Trabasso, T., van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12(1), 1–25. https://doi.org/10.1080/01638538909544717

Verhoeven, L., & Strömqvist, S. (Eds.). (2001). *Narrative development in a multilingual context* (Studies in Bilingualism 23). John Benjamins. https://doi.org/10.1075/sibil.23

Walters, J., Armon-Lotem, S., Altman, C., Topaj, N., & Gagarina, N. (2014). Language proficiency and social identity in Russian-Hebrew and Russian-German preschool children. In R. K. Silbereisen, P. F. Titzmann, & Y. Shavit (Eds.), *The challenges of diaspora migration: Interdisciplinary perspectives on Israel and Germany*. Ashgate.

Warnick, B., & Manusov, V. (2000). The organization of justificatory discourse in interaction: A comparison within and across cultures. *Argumentation*, 14, 381–404. https://doi.org/10.1023/A:1007832910904

Westby, C. E. (1995). Culture and literacy: Frameworks for understanding. *Topics in Language Disorders*, 16(1), 50–66. https://doi.org/10.1097/00011363-199511000-00005

Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts &A. Kamhi (Eds.), *Language and reading disabilities* (2nd ed., pp. 157–232). Allyn & Bacon.

Westby, C. E. (2012). Perspectives on assessing and improving reading comprehension. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities* (3rd ed., pp. 163–225). Pearson.