Applied Psycholinguistics **37** (2016), 91–122
doi:10.1017/S0142716415000430

Narratives of Russian–German preschool and primary school bilinguals: *Rasskaz* and *Erzaehlung*

NATALIA GAGARINA
Center for General Linguistics (ZAS) Berlin

ADDRESS FOR CORRESPONDENCE
Natalia Gagarina, Center for General Linguistics (ZAS), Schuetzenstrasse 18, Berlin 10117, Germany. E-mail: gagarina@zas.gwz-berlin.de

ABSTRACT
The goal of this study was to trace the dual language development of the narrative macrostructure in three age groups of Russian–German bilingual children and to compare the performance of simultaneous and sequential bilinguals. Fine-grained analyses of macrostructure included three components: story structure, story complexity, and internal state terms. Oral narratives were elicited via the Multilingual Assessment Instrument for Narratives. Fifty-eight Russian–German speaking bilingual children from three age groups participated: preschoolers (mean age = 45 months) and elementary school pupils (mean age first grade = 84 months, mean age third grade = 111 months); and there were 34 simultaneous and 24 sequential bilinguals. The results showed significant improvement for all three components of macrostructure between the preschool and first-grade period. Additional significant development from first to third graders was found only for story complexity in Russian. This is explained by the Russian curriculum explicitly teaching narrative skills during early literacy training. In the two older groups, simultaneous bilinguals showed advantages over sequential bilinguals, for story complexity only. This finding suggests considering bilingual type when evaluating narrative skills of bilinguals. The results indicate cross-language association of only some components of narrative score across languages. The findings support the examination of various constituents of macrostructure when evaluating its development as well as the progression of narrative skills.

About 30 years ago Benjamin (1985) claimed the death of narration in the broad sense: "we do not seem to generate experiences anymore that are worthy to be narrated. [. . .] A new form of report has replaced the old art of narrating, and this new form is the 'information'" (as cited in Bamberg, 1987, p. 3). Yet it seems that narratives have not died; rather they are flourishing and are the focus of numerous studies and various interdisciplinary investigations, for example, in clinical linguistics (Heilmann, Miller, Nockerts, & Dunaway, 2010; Ringmann & Siegmüller, 2013), in bilingualism and specific language impairment (Iluz-Cohen & Walters, 2011), in philosophy (Fay, 1996), and in cultural studies and literature (Nash, 1994; Newton, 1997; Polkinghorne, 1988). The ability to narrate is still considered to be the cornerstone of communication; it has an impact on social

abilities and on adequate performing in society (Bliss, McCabe, & Miranda, 1998; McCabe, 1996). Narrative skills are predictive of children's literacy and reading abilities and are crucial for later school success (Bishop & Edmundson, 1987; Bliss et al., 1998; Gutiérrez-Clellen, 2002; Hayward & Schneider, 2000; McCabe, 1996; McCabe & Rollins, 1994; Norris & Bruning, 1988; Swanson, Fey, Mills, & Hood, 2005; Torrance & Olson, 1984; Wallach, 2008). Furthermore, comparative research on narrative development has been intensively carried out in the last two decades in order to evaluate children's storytelling skills, to diagnose specific language impairment (Hayward & Schneider, 2000; Ringmann & Siegmüller, 2013; Schneider, Hayward, & Dube, 2006; Skerra, Adani, & Gagarina, 2012), and to implement narratives as a language assessment instrument (e.g., Renfrew Bus Story; Glasgow & Cowley, 1994; Test of Narrative Language; Gillam & Pearson, 2004; Edmonton Narrative Norms Instrument; Schneider, Dubé, & Hayward, 2005; and Hamburger Verfahren zur Analyse des Sprachstands Fünfjähriger; Reich & Roth, 2004). Taking into account the underdevelopment of the standardized assessment procedures for bilinguals, narratives have a strong potential, because they are less bilingually biased (Paradis, Genesse, & Crago, 2010) and provide an opportunity to examine not only the specific language skills but also the more general (cognitive, social, and pragmatic) skills of children (Liles, 1993). Narrative data are usually generated in a natural context and therefore can be seen as one of the most ecologically valid methods of investigating children's linguistic competence, as opposed to many standardized tests, which address only specific linguistic features (Botting, 2002). Thus, narrative production data can be seen as one of the most efficient ways to examine children's linguistic skills as well as their storytelling and cognitive abilities.

In order for a narrative to be understandable and appropriately perceived, it should be well formed and properly organized at two distinct, but interconnected levels: macro- and microstructure (Liles, Duffy, Merritt, & Purcell, 1995). Macrostructure refers to the higher order hierarchical organization of the narrative text such as episodic structure and story grammar components (Heilmann et al., 2010), while microstructure is represented by the internal, or language-specific, linguistic units used for the construction of coherent discourse, such as noun phrases, pronouns, and connectives. The present contribution focuses on the macrostructure of elicited narratives in both languages of bilingual children.

MACROSTRUCTURE AND INTERNAL STATE TERMS (ISTs)

Macrostructure is traditionally described as the global hierarchical organization of a text and often refers to the *story grammar model* (Mandler, 1979; Stein & Glenn, 1979). According to Stein and Glenn (1979), story grammar includes two main constituents: the setting, which introduces the characters and describes the nonlinguistic context of a story; and the episode, which consists of further components including an initiating event (an occurrence, an activity of a protagonist, etc., that triggers a response in protagonists), internal responses (the internal state of the protagonist as a response to the initiating event), internal plans (the strategy of the protagonist to reach the goal), attempts (different actions of the protagonist

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

to reach the goal), direct consequences (the [non]success of the protagonist in achieving the goal), and reactions (the protagonists' internal feelings; Peterson & McCabe, 1991; Stein & Glenn, 1979). The core components of this schema or episode constituting the story plot are the *goal* of the protagonist, the *attempt* to reach the goal, and finally, the *outcome* of the protagonist's actions.

Depending on which core parts of the episode are realized, narratives can reach various levels of complexity (according to Westby's, 2005, binary decision tree): (a) sequences with no generated goal statement; (b) incomplete episodes, which include a goal statement but lack a complete goal–attempt–outcome (GAO) structure due to omission of an attempt or outcome; and (c) the most complex and complete episodes, which include all three GAO components (McCabe & Peterson, 1984; cf. Trabasso & Nickels, 1992). The GAO represents the highest level of story complexity and is an indicator of an ability to produce a coherent story. In sum, the analysis of the number of episode components (i.e., story structure) provides a quantitative estimate of a narrative's macrostructure, and the combination of the episode components (i.e., story complexity) allows for a more in-depth qualitative evaluation. Basically, a quantitative score is based on how many story structure components the child produces when telling a story, while the qualitative complexity estimate considers the co-occurrence of goals, attempts, and outcomes within an episode.

Such an approach provides a detailed evaluation of narrative macrostructure, because it embraces both the quantity (an overall sum of episode components) and the quality (levels of complexity measured as a combination of episode components) scores. The quality score might be less language dependent; technically, one can produce fewer components of story structure, but these might be core components involving goals or even complete GAO episodes, the highest level of complexity.

Another crucial part in the evaluation of narrative macrostructure is mental state language, which is grounded in theory of mind and is realized through ISTs. ISTs provide information about children's abilities to build inferences (Nippold, Ward-Lonergan, & Fanning, 2005; Westby, 2005). Thus, the skill to produce a well-formed narrative can be said to be properly assessed if the three above-mentioned narrative constituents are evaluated.

NARRATIVE SKILLS IN DUAL LANGUAGE LEARNERS: DEVELOPMENTAL AND CROSS-LANGUAGE RELATIONSHIPS

While the development of story structure in monolingual children has been the subject of numerous studies since the midnineties (e.g., Berman & Slobin, 1994; Peterson & McCabe, 1991), narratives of bilingual children have received attention only recently (Fiestas & Peña, 2004; Gagarina et al., 2015; Gutiérrez-Clellen, Simon-Cerejido, & Wagner, 2008; Iluz-Cohen & Walters, 2011; Pearson, 2002; Uccelli & Páez, 2007). Our knowledge of the development of macrostructure and its manifestation in the two languages of bilinguals is still inconclusive with regard to children ranging in age from kindergarten (5-year-olds) through the fifth grade (10-year-olds). Moreover, studies include a restricted number of language pairs (e.g., Spanish–English; Fiestas & Peña, 2004; Gutiérrez-Clellen et al., 2008;

Montanari, 2004, a multiple case study; Pearson, 2002; Uccelli & Páez, 2007), English–Hebrew and Mandarin–English and Cantonese–English (one study each). The relevant studies are dealt with below.

Development of macrostructure in monolinguals and bilinguals

Macrostructure in monolinguals has been extensively studied in four child age groups (3-, 4-, 5, and 9-year-olds) and in adults in 12 different languages (Berman & Slobin, 1994; Trabasso & Rodkin, 1994). Berman and Slobin (1994) used the frog stories (Mayer, 1969) to trace the developmental sequence of episodic components following the story grammar model of Mandler (1979). They found that, for example, only 50% of the 4-year-olds but over 90% of 9-year-olds used mental states as initiating events. In addition, they stated that "four-year-olds produced main plot components five times as often as three-year-olds . . . nine-year-olds do so around twice as often as five year-olds" (p. 49). Similarly, Trabasso and Rodkin (1994) analyzed the realization of GAOs in English narratives and pinpointed the sharpest increase between ages 4 and 5, with slower development upward to age 9. In particular, the authors claimed that from ages 3 to 5 children moved from isolated naming of particular elements of episodic structure to the understanding and producing of GAOs, thus showing the ability to produce the highest level of macrostructure complexity by age 5 (for a description of *striking development* of macrostructure in preschool children, see also Trabasso & Nickels, 1992). Kemper (1984) reported similar findings for fantasy narratives in monolingual children between the ages of 2 and 10. She gave examples of a *true* narrative produced by children at age 5 and specified that "[b]etween the ages of 2 and 5, children pass from telling stories that are primarily heaps of unrelated events to telling true narratives that are focused around a climax" (Kemper 1984, p. 105); she further added that narrative structure is still developing beyond 5.

Few studies to our knowledge have traced the narrative development in both languages of typically developing bilinguals or compared the (full) age range dealt with in the present study (preschool to third grade). Out of these, only some studies have applied a method analogous to the one used in the present study to evaluate narrative scores in bilingual children of approximately similar ages. Uccelli and Páez (2007) performed narrative elicitation tasks (with a wordless picture book Frog story; Mayer, 1969) with Spanish–English kindergartners and first graders (i.e., 5- and 6-year-old children) and reported significant improvement with age in development for story scores (comparable to the two analyses described above) in both languages. They adapted the Pearson's (2002) composite narrative score scheme evaluating the elements of story structure, such as initiating events, attempts, goals, and so on, as well as hierarchal story sequence such as perspective, including metacognitive elements and temporal linking. Pearson (2002) herself also found significant growth on the same measures in Spanish–English bilinguals from second to fifth grade in Frog story narratives, while the difference in performance of bilinguals and monolinguals was only "quite small" (Pearson, 2002, p. 148).

In sum, whereas in the developmental studies on monolingual children the macrostructure of elicited children's narratives has been shown to develop

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

considerably between ages 3 and 5 and also beyond 7, its active development was reported for bilingual kindergarten and first-grade children (5- to 6-year-olds) and for the children until the fifth grade (see also Bishop & Donlan, 2005; van den Broek, 1997).

Cross-language relationship of narrative macrostructure

Pearson's (2002) investigation of the narrative skills of English–Spanish second- and fifth-grade bilinguals made use of a composite narrative score that was found to be correlated in the two languages. Pearson (2002) also examined the children's microstructure abilities, analyzing vocabulary, morphosyntax, and fluency, and finding few or no cross-language correlations on these measures. On the basis of these findings, she suggested that "the narrative/discourse elements (Story Score) showed carry-over across languages" but not the microstructure measures (Pearson, 2002, p. 149). Fiestas & Peña (2004) reported similar results for elicited narratives in the same pair of languages in kindergarten and elementary school children (age range = 4 years, 0 months [4;0] to 6;11), but their evaluation of narrative skills was somewhat different. Macrostructure was coded for inclusion of story grammar elements produced (i.e., setting, initiating event, internal response, plan, attempt, consequence, and ending; following Stein & Glenn, 1979), and overall complexity was calculated for each story. While narrative complexity was relatively equivalent in both languages, thus suggesting cross language transfer, initiating events and attempts appeared more frequently in Spanish narratives, and consequences in English narratives (Fiestas & Pena, 2004).

Cross-language association of narrative skills was the aim of the Uccelli and Páez (2007) study. Following a narrative elicitation task with kindergartners and first graders (5- and 6-year-old children, respectively), the authors evaluated narrative story scores based on Pearson's (2002) scheme. The story score included "coding and scores for three subcomponents: story elements, story sequence, and perspective" (Uccelli & Páez, 2007, p. 6) and was used to evaluate the child's production of a coherent flow of events and their components. Consistent with the previous studies, a positive moderate association of Spanish and English narrative skills (i.e., macrostructure) in both age groups was found. The authors underlined the importance of the contribution of Spanish narrative skill to English narrative score.

Similar findings were reported by Iluz-Cohen and Walters (2012) for English–Hebrew sequential bilingual preschoolers, eight typically developing children and nine with language impairment, with mean ages of 5;11 and 6;1, respectively. The authors targeted macro- and microstructure of two familiar stories elicited in both languages, *Jungle Book* in English and *Goldilocks and the Three Bears* in Hebrew. Their results for macrostructure and microstructure supported their main hypothesis that "story structure should be invariant across a bilingual child's two languages and should lead readily to cross-linguistic association, while lexical and morphosyntactic abilities should be more language specific and less predisposed to transfer" (Iluz-Cohen & Walters, 2011, p. 3).

In addition, Gagarina et al. (2015) in the Language Impairment Testing in Multilingual Settings (LITMUS)-MAIN study report for 302 bilingual children in the

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

age range of 3;6 to 9;1 (with 14 different language pairs, such as second language–first language [L2-L1] Afrikaans–English, Dutch–Turkish, English–Hebrew/–Polish, Finnish–Russian/–Swedish, German–Russian/–Turkish, Greek–Albanian, Italian–English, Lithuanian–Russian, Swedish–English/–French/–Russian) that "bilingual children score equally high on story structure in both their languages even when parents report one of the child's languages to be weaker than the other. [. . .] Similar story structure scores in both languages also provide support for the view that abilities at the macrostructure level, i.e. composition of cohesive event sequences, reflect capacities which go beyond the specifics of one language" (pp. 261–262).

Finally, a study by Jia, Yiu, Sorenson Duncan, and Paradis (2010) showed no significant differences in macrostructure between the two languages of 14 Mandarin–English and 12 Cantonese–English bilingual children ages 6;7 to 9;7. For the evaluation, they used the scoring sheet of the Edmonton Narrative Norms Instrument (Schneider et al., 2005) based on the story grammar evaluation of Stein and Glenn (1979).

All studies reviewed above reported interrelatedness of macrostructure skills across the two languages of bilingual children, measured in terms of story grammar components, whereas microstructural (e.g., lexical and morphosyntactic) components appeared not to associate across languages thus staying language specific.

BILINGUAL TYPE: SIMULTANEOUS VERSUS SEQUENTIAL

Simultaneous and sequential bilinguals differ in age of L2 onset, although there is little agreement as to what age the onset of exposure should begin in order for a child to be defined as a sequential bilingual. The age of L2 onset for sequential bilinguals has been variably identified as 4 to 8 years by Meisel (2009), 4 to 6 years by Montrul (2008), and 3 to 6 years by Paradis (2008). In the present study, sequential bilinguals are defined, following Ruberg (2013), as children with L2 onset later than 24 months. Little is known about variation in the development of linguistic skills in these two types of bilinguals, because their dual languages are rarely compared in the same study; however, differences were found between the brains of simultaneous and sequential bilinguals, the latter defined as the age of onset after the age of 3 (Mohades et al., 2012).

The age of onset has been intensively discussed within two age brackets: around puberty and from birth to age 4. Around puberty, studies on L2 acquisition investigated attainment within the critical period hypothesis, and for younger children, the differences between simultaneous and sequential bilingualism were dealt with in age ranges from birth to approximately age 4 (for a review, see Li, 2013; Muñoz & Singleton, 2011).

Studies on language development in simultaneous bilinguals emphasize the similarity in the progression of two languages of a bilingual child to monolingual acquisition (e.g., see Hakuta, 1986, for summary of these studies; and Döpke, 2000, for cross-linguistic structures).

The role of age of onset (together with the chronological age and length of exposure) in L2 was investigated in Armon-Lotem, Walters, and Gagarina (2011). In their comparative study, 65 Russian–German and 78 Russian–Israeli children

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

(divided in three age groups: 4-, 5-, and 6-year-olds) with a mean length of L2 exposure of 37 months were tested six to seven times in both languages with a battery of language tasks including standardized tests for L2s. All three age groups were below but within one standard deviation of the monolingual means. For an in-depth analysis as a function of length of exposure, participants were divided into three groups in terms of short (<25 months), middle (25–48 months), and long length of exposure (>48 months). The results showed that all groups, even children with 4 and more years of exposure, did not reach the monolingual mean. For an analysis of age of onset, children were divided into three groups in the Israeli cohort: earlier age of onset (<25 months), intermediate age of onset (25–36 months), and later age of onset (>36 months). All in all, age of onset was shown to have significant negative correlations with performance on the L2 Hebrew tasks. Children with age of L2 Hebrew onset before the age of 2 "performed within the monolingual norms (in the range of +/- one SD) both in terms of the total score on the standardized test and in terms of its subsets. On the other hand, those who began L2 acquisition after the age of 3 were more than 1.5 SDs below the monolingual norms for the total score and for the vocabulary score (even though the mean *length of exposure* was 21 months)" (Armon-Lotem et al., 2011, p. 301). In the German cohort, in which the absolute majority of children were sequential bilinguals, with the age of onset after age 2, no correlation of age of L2 onset and performance in L2 tests was found. In contrast to the Russian–Hebrew cohort, it was chronological age that significantly correlated with all tests.

However, to my knowledge, none of the studies investigated dual language development of either simultaneous or sequential bilinguals with the emphasis on the possible peculiarities in their productive narrative skills depending on the type of bilingualism. The present study thus attempts to start filling in this gap by comparing story structure, story complexity, and IST primary school children who are sequential versus simultaneous bilinguals. The results might have societal relevance: if sequential bilinguals with age of onset 24 months will have (sustained) disadvantages either in L2 German or even in L1 Russian narratives in comparison with their simultaneous peers, this will speak for the necessity of even earlier focused support of L2, namely, prior to 24 months.

THE STUDY: GOALS AND HYPOTHESES

In the present study, the narrative macrostructure of children in three age groups (preschool, first grade, and third grade) and two bilingual groups (simultaneous and sequential) was investigated. The study had two primary goals. The first goal of this study was to trace the development of three components of macrostructure (story structure, story complexity, and ISTs) in bilingual children's home first language (L1 Russian) and environment language (German, L2). Given the results of previous studies showing that narrative skills develop considerably around age 5 but also beyond and that children at this age produce GAOs, I predicted developmental growth in all three components of macrostructure for both languages, with the greatest differences between the preschool children and the primary school groups and a similar developmental trajectory across dual languages based on

previous findings (Iluz-Cohen & Walters, 2011; Pearson, 2002; Simon-Cereijido & Gutiérrez-Clellen, 2009).

The second goal was to compare narrative skills of simultaneous and sequential bilinguals in primary school in order to examine possible variation in the macrostructure of the narratives of these two bilingual groups. Sequential bilinguals were expected to score below simultaneous bilinguals in L2 German narratives on story structure and ISTs throughout the age range studied. This prediction is grounded (indirectly) on the evidence, showing that age of onset influences the attainment of L2 and the maintenance of L1 (Muñoz & Singleton, 2011), that sequential bilinguals are behind the monolingual norm on language tests even after 4 years of exposure to L2 (Armon-Lotem et al., 2011; Gagarina et al., 2014). Story complexity is expected to be similar for simultaneous and sequential bilinguals and across their two languages, because this macrostructure component is associated with the more general (cognitive) knowledge of constructing well-formed discourse. The study thus fills a gap in investigating narrative skills between sequential and simultaneous bilinguals because there are hardly any studies comparing these two populations.

Another final, theoretical goal was to provide a more differentiated "picture" of claims of macrostructure universality and to explicate macrostructure's granularity. I do so by tracing the development and by comparing all three components of macrostructure separately.

METHOD

Participants

Fifty-eight bilingual children participated; 21 preschoolers, 15 first graders, and 22 third graders. All were conversant in Russian as their first language and in the Berlin variant of standard German as their societal language. All had at least one parent who was a native speaker of Russian and a first-generation immigrant. The children in the preschool group were recruited among preschool children from different Berlin districts; all attended full-time monolingual German-speaking kindergartens (similar to a nursery school in the US system) with no more than half the preschool students from immigrant families. The main language of communication in these kindergartens is German. The children from the two older groups attended an elementary state bilingual school in the center of Berlin with a 5-day school week and an 8-hr school day. Instruction in this school is divided equally between the two languages: half the subjects are taught in German by teachers who are (mostly) native speakers of German and half are taught in Russian by native speakers of Russian. The language of communication during recess and on the schoolyard is predominantly German.

For both the preschool and the elementary school children, the parents' socioeconomic status ranged from low income to upper middle class, including university educators/professors. The mean social status index did not differ significantly for the first graders (mean social status index = 16.45, $SD = 1.18$) and third graders (mean social status index = 16.14, $SD = 1.07$), but was significantly lower for the preschoolers (mean social status index = 12.42, $SD = 2.03$). The social status

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

Table 1. *Number and means (standard deviations) for chronological age and age of onset of German for the three groups of bilingual children and the subdivision into simultaneous and sequential groups*

Three Groups of Bilingual Children ($N = 58$)	Preschoolers ($N = 21$)	1st Graders ($N = 15$)	3rd Graders ($N = 22$)
Chronological age (months)	45 (5.9)	84 (3.4)	111 (8.2)
Age range (months)	31–52	77–89	95–126
Age of onset of German (months)	26 (7.0)	27 (14.1)	12 (17.4)
Age range of onset of German (months)	10–39	9–50	0–50
Simultaneous ($N = 32$)	$N = 7$	$N = 8$	$N = 17$
Mean chronological age (months)	43 (6.4)	84 (4.0)	112 (8.3)
Age range (months)	35–49	77–89	95–126
Mean age of onset of German (months)	19 (4.8)	16 (5.5)	4 (7.6)
Age range of onset of German (months)	10–24	9–22	0–24
Sequential ($N = 26$)	$N = 14$	$N = 7$	$N = 5$
Mean chronological age (months)	46 (5.7)	85 (2.6)	107 (7.5)
Age range (months)	31–52	81–89	96–115
Mean age of onset of German (months)	30 (4.4)	39 (9.1)	42 (10.1)
Age range of onset of German (months)	25–39	26–50	25–50

index was calculated according to the guidelines of the Berlin Senate Department for Health and Social Affairs (Bettge & Oberwoehrmann, 2012) and included three components: school-graduation qualification (in German *Schulabschluss*), professional education (in German *Berufsausbildung*), and employment status (in German *Erwerbsstatus*). These were weighted equally and calculated for both parents: 0 to 3 points per parent. If the data for one parent was missing, the scores of the other parent for a given feature were doubled.

Prior to testing, all parents provided written consent agreeing to participation in the study. In addition, parents completed a questionnaire (i.e., Russian Language Proficiency Test for Multilingual Children; Gagarina, Klassert & Topaj, 2010) providing background information about the child, developmental milestones, quantity and quality of language input, social status, medical history, and so forth. On the basis of these reports, 32 children were identified as simultaneous bilinguals and 26 as sequential bilinguals (exposed to L2 German after 24 months).

All children were also given IQ tests: the preschool group took the Snijders–Oomen Nonverbal Intelligence Test (Tellegen, Laros, & Petermann, 2007), and the two older groups took the Raven's Coloured Progressive Matrices (Bulheller & Häcker, 2002). On the basis of the parental questionnaire, IQ tests and conversations with the teachers, children with deficits in hearing, neurological development, and cognitive development, or with articulation problems were excluded.

The upper part of Table 1 displays the three age groups: their mean ages, age range, and age of onset to L2. The lower part of Table 1 shows the same numbers given separately for simultaneous and sequential bilinguals.

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

Language (proficiency) tests

Home language (Russian) skills were tested in the preschool and first-grade groups with the Russian Language Proficiency Test for Multilingual Children (Gagarina et al., 2010). The test evaluates productive and receptive lexicon for verbs and nouns, production of morphological inflections on verbs (first- and second-person singular present), production of case on nouns (accusative and dative), and comprehension of sentence level grammatical constructions, which are relevant in the acquisition of Russian, like devočka gladit ruku perom *girl-NOM stroke-IMPV.PRES.3SG hand-ACC feather-INST* "The girl is stroking the hand with the feather" or devočku tolkaet malčik *girl-ACC push-IMPV.PRES.3SG boy-NOM* "The boy is pushing the girl." This test is performed as a picture-selection task based on the Test of Reception of Grammar (Bishop, 1989; Fox, 2006). Preliminary norms (unpublished) are based on data from almost 400 bilingual children ages 3–9.

Language skills in German were assessed with two different tests, one for preschool children and the other for the first-grade school group. Preschoolers were assessed with Linguistische Sprachstandserhebung—Deutsch als Zweitsprache (Schulz & Tracy, 2011), a test normed for sequential bilingual children from 3 to 6 years. It consists of seven subtests and covers production and comprehension of basic morphological, syntactic, and semantic structures. Subtests for lexicon (production and perception) and for syntax (perception) were administered. Language skills in the first-grade school group were assessed with the WWT 6–10 (Glueck, 2011). The test is normed for monolinguals aged 6–10, and it targets the production and comprehension of semantic–lexical abilities. For the third graders, the LITMUS Sentence Repetition Test (for Russian: Meir, Armon-Lotem, & Walters, 2015; for German: Hamann, Chilla, Ruigendijk, & Abed Ibrahim, 2013) was administered in both languages; the evaluation of the results showed no significant differences between the two languages. In addition, between ages 5 and 6, all the school children were administered language tests in Russian and German for admission to the school. On the normed and standardized tests, all children performed within 1.25 *SD* of the mean.

Narrative elicitation, transcription, and coding

Narratives were elicited with the LITMUS-MAIN instrument for the assessment of narrative abilities of bilingual children. The instrument allows elicitation of multiple languages from the same child via "telling" (story generation) or "retelling" and includes a set of comprehension questions focused on goals and mental states. For the present study, a model story and subsequent telling were administered. "Baby Goats" (based on Hickmann, 2003) and "Baby Birds" (developed from Guelzow & Gagarina, 2007) were used for story telling and Cat and Dog stories were used as model stories (see Appendix A for an example story). The full set of guidelines for administration and narrative elicitation, story scripts and picture stimuli, scoring protocols, comprehension questions, and background questionnaire are available in 26 languages from http://www.zas.gwz-berlin.de/zaspil.html.

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

A native speaker of either German or Russian or a bilingual investigator administered the LITMUS-MAIN to all bilingual children in a monolingual mode with an interval of at least 4 days between languages. For example, one child encountered the Cat story in the model story format and the Baby Goats story in the telling format, both in German; then she encountered the Dog story in the model story format and Baby Birds in the telling format, both in Russian. The order of presentation was counterbalanced for storytelling format (model/telling) and for language (Russian/German).

At the beginning of the procedure, a set of warm-up questions is asked (e.g., *Tell me about your friend*, *What toys do you like*, *What is your favorite book or film*, etc.). After the warm-up phase, the experimenter places three colored envelopes on the table and asks the child to choose one, saying: *There is a different story in each envelope. Choose one and then I will tell you the story*. The child chooses an envelope and removes a set of six pictures in the form of an accordion. The experimenter unfolds the pictures so the entire sequence is visible only to the child (in order to provide nonshared attention). After the child has looked through the entire sequence of pictures, the experimenter asks: *Are you ready? I am going to tell you the story and then I will ask you some questions*. The story is told and the 10 comprehension questions are asked. Following the first story, the experimenter moves to the second part of the assessment. She places another three colored envelopes on the table and asks the child to choose one, saying: *There is a different story in each envelope. Choose one*. *Now I want you to tell me the story. Look at the pictures and try to tell the best story you can.* The experimenter helps the child, if necessary, to unfold the pictures in order to get an overview of the entire set. Then the story is refolded, so that the child sees only the first two pictures, and she begins telling the story. If the child is reluctant to begin, she is prompted as follows: *Tell me the story* (the experimenter points to a picture). When the child has finished telling the story in the first two pictures, the experimenter unfolds the next two pictures (so that four pictures are now visible). The experimenter then repeats the process until the child has reached the end of the story, and finally she asks 10 comprehension questions (for the allowable prompts, see Gagarina et al., 2012, p. 112).

The narratives were transcribed verbatim by native speakers of the two languages using Codes for Human Analysis of Transcripts conventions (MacWhinney, 2000) and morphologically tagged in Russian with the MORCOMM program (Gagarina, Voeikova, & Gruzincev, 2003). The transcription for the Russian stories was performed in St. Petersburg by an expert with long-time experience in working with elicited narratives. The transcripts were then checked by the present author. For the German narratives, the transcribers and coders were trained and their transcripts were then double checked by a third person for accuracy. The procedure included transcribing one third of audio files per group and performing a Cohen kappa reliability score with Computerized Language Analysis; the agreement on the transcribed words was over 80%. Coding and analyses of story structure, story complexity, and ISTs were performed by native speakers of either German or Russian according to the protocols of LITMUS-MAIN, in which the number of possible answers was presented as exhaustively as possible. If they were not sure as to a denotation, a group of two or three researchers discussed the case until a general consensus was reached.

Narrative measures

Three components of macrostructure were analyzed: *story structure* (a quantitative measure, with a maximal score of 17 points), *story complexity* (a qualitative measure, with a maximal score of 3 indicating the production of three complete GAO episodes), and *ISTs* (the number and percentage of IST terms, reflecting the narrator's production of internal state words).

In terms of story structure, the presence of setting information (max 2 points: 1 for time, 1 for place) and the number of goals, attempts, outcomes, ISTs produced as an initiating event, and ISTs produced as a reaction were tallied. Because each set of pictures depicted a story that contained three episodes, it was possible to produce three of each component except the setting (see Appendix A, Picture 1). Repetitions of the same component were excluded from this calculation. Thus, the maximal number of points possible for a story was 17 (2 for setting; 9 for each of three Gs, As, and Os; and 6 points for ISTs as either initiating events or reactions).

The calculation of narrative complexity was based on Westby's (2005) binary decision tree. The highest level of structural complexity achieved by a child in his/her story productions was identified: *no* elements of macrostructure (not A, O, or G; score of 0); A, O, or the combination of both (score of 1); G, GA, or GO (score of 2); or the full GAO sequence (score of 3). The proportion of children who achieved each level of complexity (0–3) was tallied for each group. In addition, the number of times a child produced a full GAO sequence in their narrative was calculated (max number = 3).

Finally, all IST tokens in a narrative (outside of mazes) were counted, and the percentage of ISTs out of all word tokens was calculated. ISTs included words denoting different internal states: *perceptual state terms* (e.g., see, hear, feel, and smell), *physiological state terms* (e.g., thirsty, hungry, and tired), *consciousness terms* (e.g., awake and asleep), *emotion terms* (e.g., sad, happy, and angry); *mental verbs* (e.g., want, think, and know), and *linguistic verbs* (e.g., say, call, and ask). This measure included all produced tokens, and it was not restricted to ISTs as initiating and reaction events only.

RESULTS

The results are presented in the following steps: the comparison of the story structure (quantitative analysis), story complexity (qualitative analysis), and ISTs for the three age groups in both languages is given. As far as cross-language comparison is concerned, for each of the three measures, correlations between L1 Russian and L2 German are provided. In the second part of the study, simultaneous and sequential bilinguals are compared on the same measures. Parametric tests are used unless their assumptions are violated, in which case nonparametric tests are used.

Age and language differences

The mean scores for the number of story structure elements produced (maximum 17) for the three age groups and two languages are given in Table 2. For L2 German a consistent increase with age was observed, whereas for L1 Russian a

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

Table 2. *Story structure: Mean (standard deviations) scores, interlanguage correlations, and p values for three age groups and two languages*

Three Groups of Bilingual Children ($N = 58$)	Preschoolers ($N = 21$)	1st Graders ($N = 15$)	3rd Graders ($N = 22$)
Mean score in Russian (*SD*)	3.53 (2.48)	7.47 (2.07)	6.86 (2.17)
Mean score in German (*SD*)	3.38 (2.16)	8.47 (1.51)	9.73 (2.19)
Correlation Between Languages			
r	.56	.52	.26
p	.01	.05	.24

Note: The story structure score summarizes single components: internal state terms as initiating events and reactions, goals, attempts, and outcomes for each of the three episodes as well (maximum 17 points).

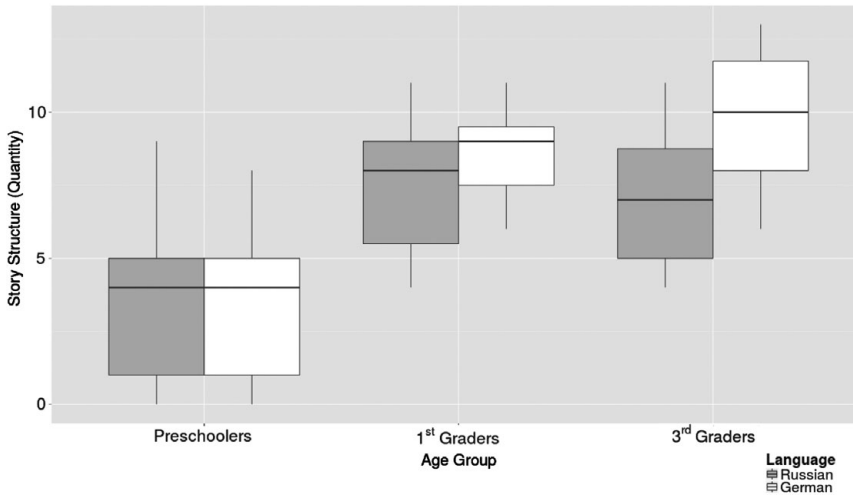


Figure 1. Story structure in Russian and German in three age groups. The story structure score summarizes single components for each of the three episodes: setting including 1 point for each time and place reference, internal state items as initiating events, and reactions, goals, attempts, and outcomes (maximum 17 points).

slight decrease from first to third grade was documented. One-way analyses of variance (ANOVAs) done separately for each language revealed significant main effects for age group for both Russian, F (2, 53) = 16.18, $p < .001$, and German, F (2, 55) = 57.47, $p < .001$, narratives. For both Russian and German, post hoc Tukey tests showed significant differences between the preschool and the first- and third-grade groups ($p < .001$), but no significant differences between the two elementary school groups (see Figure 1 for visual representation).

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

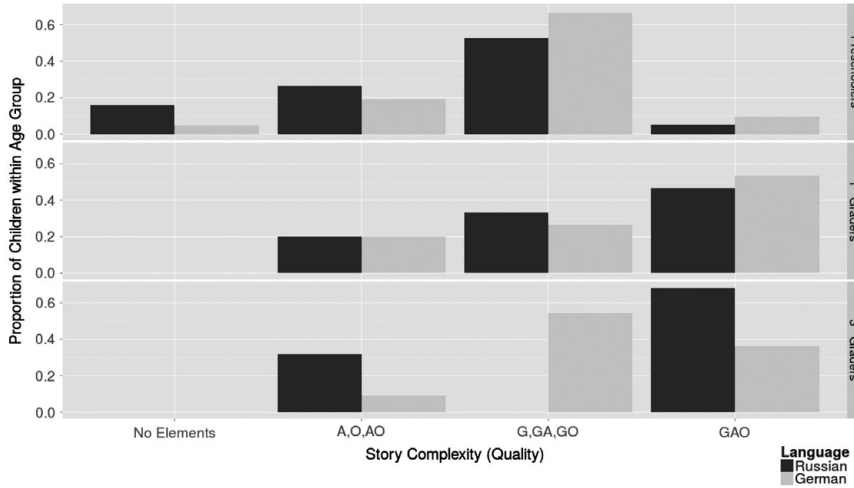


Figure 2. Story complexity in Russian and German in three age groups. G, Goal; A, attempt; O, outcome. Story complexity reflects different levels of a narrative: sequences with no generated goal statement (A, O, AO); incomplete episodes, which include a goal statement but lack a complete GAO structure due to omission of an attempt or outcome (G, GA, GO); and the most complex and complete episodes including all three GAO components.

Correlations assessed the relationship between the two languages on the quantitative measure of story structure for each age group (see Table 2). The strongest correlations were found for the preschool and first-grade groups, and these correlations were significant. In third grade, the correlation was weaker and not significant.

For story complexity, the proportion of children producing episodes containing various story elements as their highest level of story complexity in each age group is shown in Figure 2. The first category (no elements of macrostructure) exists only in preschoolers. The two intermediate levels (sequences with no generated goal statement and incomplete episodes, which include a goal statement but lack a complete GAO) are often present in preschoolers but less so in first and third graders. The highest level of story complexity (complete episodes including all three GAO components) is rarely present in preschoolers in either language, but is produced more often by older children (i.e., about 50% of first graders produced GAOs). A separate analysis of the highest complexity level (GAO), using Fisher's exact test for count data, showed a significant change across the three age groups in the proportion of children producing a GAO in Russian and in German ($p < .001$ and $p = .014$, respectively). Post hoc analyses revealed significant differences between preschoolers and first graders in both languages (Russian: $p = .02$; German: $p = .02$) as well as between preschoolers and third graders in Russian ($p = .001$), but not in German ($p = .14$). In both Russian and German, first and third graders did not differ significantly from each other (Russian: $p = .31$; German: $p = .34$). A comparison of the languages in each age

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

group revealed no differences for preschoolers and first graders. There was a trend for a lower proportion of third graders to produce GAOs in German than in Russian ($\chi^2 = 3.28$, $df = 1$, $p = .07$).

A Kruskal–Wallis test was used to analyze age group differences on the complexity scores. Significant differences were obtained between the age groups in both Russian and German ($\chi^2 = 10.94$, $df = 2$, $p < .002$ and $\chi^2 = 12.51$, $df = 2$, $p < .002$, respectively). Post hoc analyses (pairwise comparisons using Wilcoxon rank sum test) revealed significant differences between preschoolers and first graders in both languages (Russian: $p = .040$; German: $p = .014$) as well as between preschoolers and third graders (Russian: $p = .005$; German: $p = .004$). However, in both Russian and German, first and third graders did not differ significantly from each other (Russian: $p = .548$; German: $p = .606$).

The interlanguage correlations for the complexity scores were low for all three groups (Pearson $r = .06$, .18, and .13, for preschool, first and third grades, respectively), and the calculation of significance ($p = .74$, .41, and .46, respectively) shows that the data might be compatible with randomness.

Table 3 shows the total length of narratives in words. One-way ANOVAs in each language showed that an increase with age was significant for Russian: F (2, 53) = 5.9, $p < .05$. Post hoc analyses revealed that the story length in preschoolers differed significantly from the story length of first graders ($p = .005$); the difference in story length between preschoolers and third graders approached significance ($p = .051$), and it was not significant between first and third graders ($p = .480$). Similarly, for German the age groups differed significantly on story length: F (2, 55) = 22.2, $p < .01$. Post hoc analyses also showed the story length in preschoolers differed significantly from both first graders and third graders (both $p < .001$), but the comparison of first graders and third graders revealed no significant difference ($p = .510$).

To answer the question of whether bilingual children in the three age groups significantly differed in the production of ISTs, the total number of IST tokens were first analyzed, and given that ISTs are text-length dependent, the percentage out of the total number of narrative word tokens was examined (Table 3). Kruskal–Wallis tests of the number of ISTs produced revealed significant differences both for Russian ($\chi^2 = 13.46$, $df = 2$, $p = .001$) and German ($\chi^2 = 31.54$, $df = 2$, $p < .001$). Post hoc analyses for Russian and German showed significant differences for preschoolers versus first graders (Russian: $p < .002$; German: $p < .001$) and preschoolers versus third graders (Russian: $p < .020$; German: $p < .001$) but not between first and third graders (Russian: $p = .300$, German: $p = .260$).

One-way ANOVAs of the proportion of ISTs to total word tokens for each language revealed no significant age differences for Russian, F (2, 52) = 2.1, $p < .13$, but a significant growth in the percentage of ISTs from preschool to first grade only in German, F (2, 55) = 12.8, $p < .01$.

An examination of individual use of ISTs in Russian and in German in the three age groups revealed that in Russian 10 preschoolers but only 4 children in third grade produced no ISTs; and the number of children producing more than 2 ISTs increased with age, although no child produced 5 or more ISTs. In German, the picture is similar for the preschoolers. Beginning with elementary school, however, a different pattern is observed: the number of children producing

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

Table 3. *Means (standard deviations) for length of stories in word tokens and internal state terms (ISTs) in tokens and in percentages (standard deviations) for three age groups and two languages*

	Preschoolers (N = 21)	1st Graders (N = 15)	3rd Graders (N = 22)
	Russian		
Story length in tokens	53.70 (35.60)	89.60 (30.00)	77.30 (28.10)
IST in tokens	0.79 (0.71)	2.07 (0.65)	1.73 (0.78)
IST (%)	1.45 (1.92)	2.54 (0.99)	2.15 (1.45)
	German		
Story length in tokens	47.30 (29.10)	106.80 (45.90)	120.90 (39.20)
IST in tokens	0.86 (1.08)	5.2 (2.91)	3.86 (1.92)
IST (%)	1.78 (2.42)	4.87 (1.37)	3.12 (1.17)
	Correlation of ISTs (%) Between Languages		
r	−.11	−.59	.37
p	.67	.02	.09

Note: ISTs include perceptual state terms, for example, *see, hear, feel, smell*; physiological state terms, for example, *thirsty, hungry, tired, sore*; consciousness terms, for example, *alive, awake, asleep*; emotion terms, for example, *sad, happy, angry, worried, disappointed*; mental verbs, for example, *want, think, know, forget, decide, believe, wonder, have/ make a plan*; linguistic verbs/verbs of saying/telling, for example, *say, call, shout, warn, ask.* The negative correlation in preschoolers and first graders is because of the following: nearly all (except for two preschoolers, who showed the different pattern) children have similarly low percentages of IST in Russian, but three preschoolers and four first graders have many ISTs, which lead to high percentages, in German, so that the negative correlation occurs.

at least 4 ISTs increased from first to third grade and 10 children (5 first graders and 5 third graders) used 5 to 13 ISTs in their narratives. Thus, language-specific behavior in IST production is observed with the more advanced development in German.

Correlations between languages on the number of ISTs were moderate (τ = 0.31, –0.25, and 0.36 for preschoolers, first and third graders, respectively) but nonsignificant (p = .19, .37, and .10). For the proportion of ISTs, only the correlation for the first graders was large (τ = –0.59) and significant (p = .019; Table 3). These findings might suggest that ISTs are more language dependent and are contingent on the language-specific lexical knowledge of children.

Simultaneous versus sequential bilinguals

In the second part of the study, the results for simultaneous and sequential bilinguals were compared. Preliminary analyses of story structure, story complexity,

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

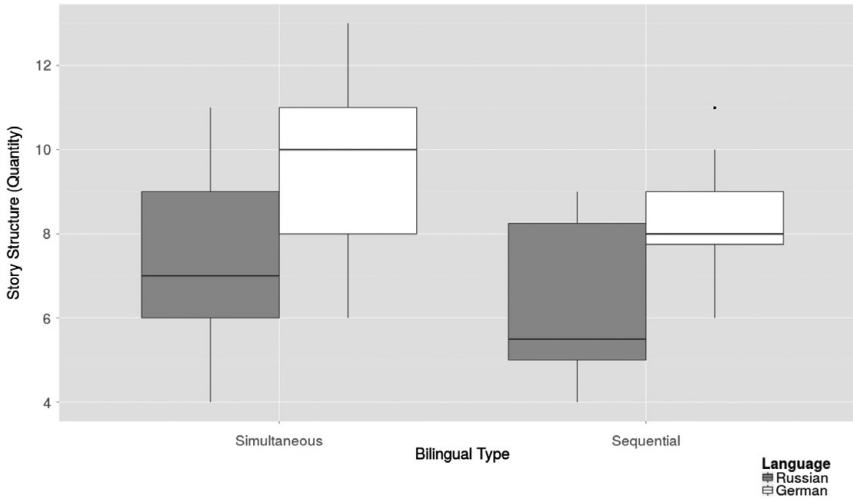


Figure 3. Story structure in Russian and German in simultaneous and sequential bilingual children. The story structure score summarizes single components for each of the three episodes: setting including 1 point for each time and place reference, internal state items as initiating events, and reactions, goals, attempts, and outcomes (maximum 17 points).

and ISTs between simultaneous and sequential preschoolers revealed no significant differences, so the following analyses were performed on data from a combined group of first- and third-grade elementary school children, which were grouped according to the bilingual type. To repeat, the onset of exposure to German for sequential bilinguals was after 24 months (following Ruberg, 2013).

For the quantitative measure of story structure (see Figure 3 and Table 4), in both languages simultaneous bilinguals scored higher than sequential bilinguals. This difference, however, as the Welch two-sample *t* tests exhibited, was significant only for German ($t = 2.385$, $df = 29.67$, $p = .02$). Correlations of story structure between languages were medium to large and not significant for simultaneous bilinguals (Kendall's rank correlation test $\tau = 0.50$, $p = .09$) but approaching significance for sequential bilinguals ($\tau = 0.46$, $p = .06$).

The proportion of children who produced each of three levels of story complexity are shown in Figure 4. The upper part of the figure depicts the results for simultaneous bilinguals, who were more likely to produce a full realization of an episode, GAO, in both languages and less likely to produce lower levels of story complexity (A, O, and AO productions). For Russian and German, the story complexity was shown by a Wilcoxon rank sum test to be significantly higher in simultaneous as compared to sequential bilinguals ($W = 216$, $p = .02$ and $W = 212.5$, $p = .04$, respectively).

An analysis of the highest complexity level, GAO production, Fisher's exact test showed that the proportion of simultaneous bilinguals who produced GAOs was significantly higher than the proportion of sequential bilinguals in Russian ($p = .036$, odds ratio = 0.204) but not in German ($p = 0.17$, odds ratio = 0.317).

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

Table 4. *Means (standard deviations) for story structure, length of stories in word tokens, and internal state terms (ISTs) in tokens and percentages (standard deviations) for simultaneous and sequential bilinguals*

	Preschoolers (N = 7/N = 14)	1st and 3rd Graders (N = 25/N = 12)
	Russian: Simultaneous/Sequential	
Story structure	3.3 (2.4)/3.7 (2.6)	7.4 (2.2)/6.5 (1.9)
Story length in tokens	58.00 (25.22)/43.79 (40.17)	85.36 (32.99)/67.50 (16.24)
IST in tokens	0.85 (0.99)/0.64 (1.10)	1.88 (1.11)/1.83 (0.79)
IST (%)	1.66 (2.04)/1.32 (2.02)	2.18 (1.43)/2.57 (1.04)
	German: Simultaneous/Sequential	
Story structure	3.0 (1.9)/3.6 (2.3)	9.7 (2.1)/8.2 (1.5)
Story length in tokens	44.00 (22.39)/43.57 (28.73)	120.84 (41.56)/99.83 (31.10)
IST in tokens	0.71 (0.88)/0.92 (1.16)	4.76 (2.79)/3.67 (1.31)
IST (%)	1.59 (2.61)/1.87 (2.50)	3.80 (1.63)/3.89 (1.42)

Note: The story structure score summarizes single components for each of the three episodes: setting including 1 point for each time and place reference, ISTs as initiating events and reactions, goals, attempts, and outcomes (maximum 17 points). The p values for the differences between simultaneous and sequential (all three age groups taken together) are the following: story structure Russian: p = .060; story structure German: p = .005; IST in percentage Russian: p = .839; IST in percentage German: p = .376.

In addition, the relationship between the number of complete GAO episodes produced per child in each language was analyzed. Kendall's rank correlation tests revealed large correlations on this measure between Russian and German for simultaneous and sequential bilinguals (τ = 0.51 and 0.69, respectively) that, however, were not significant for Russian (p = .17) and showed a trend toward significance for German (p = .07).

A Welch t test compared story length for simultaneous and sequential bilinguals. It revealed no significant differences between the two groups for German, F (1, 35) = 2.21, p = .15, or Russian, F (1, 35) = 3.27, p = .79.

Table 4 shows higher mean numbers of ISTs in both languages for simultaneous in comparison to sequential bilinguals. However, A Welch two-sample t test showed these differences were not significant. In addition, no significant differences were obtained when comparing the proportion of ISTs for German, F (1, 35) = 0.03, p = .87, or for Russian, F (1, 35) = 0.71, p = .41. Kendall's rank test showed no significant cross-language correlations either on the number of ISTs for sequential (τ = 0.22, p = .17) or simultaneous bilinguals (τ = 0.04, p = .26) or for the proportion of ISTs (sequential τ = 0.22, p = .33 and simultaneous τ = 0.04, p = .79 bilinguals).

In sum, although descriptively simultaneous bilinguals seem to show better performance on all three components of macrostructure, significance was found only for story structure for German and for story complexity for Russian.

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

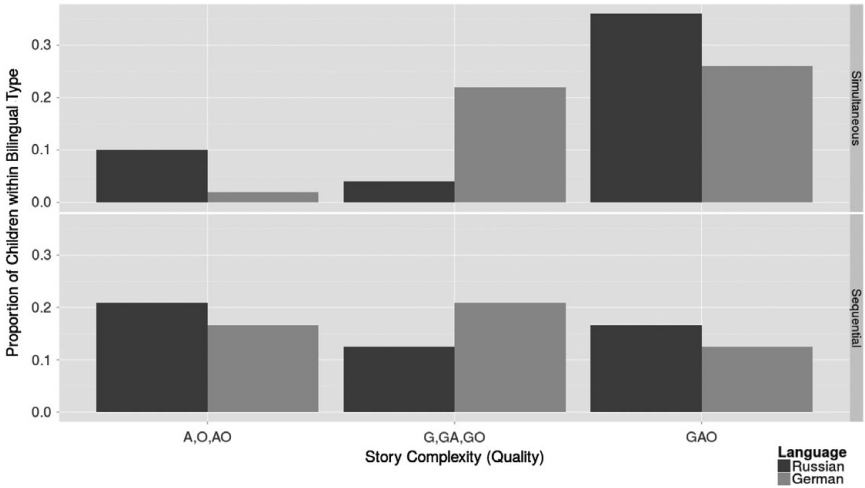


Figure 4. Story complexity in Russian and German in simultaneous and sequential bilingual children. G, Goal; A, attempt; O, outcome. Story complexity reflects different levels of a narrative: sequences with no generated goal statement (A, O, AO); incomplete episodes, which include a goal statement but lack a complete GAO structure due to omission of an attempt or outcome (G, GA, GO); and the most complex and complete episodes including all three GAO components.

DISCUSSION

The developmental patterns for story structure, story complexity, and ISTs were compared for three age groups (preschoolers, and first and third graders in primary school) and two languages (Russian and German). Data come from 58 bilingual Russian–German children from 2;6 to 10;6, divided into a preschool group of 21 children with a mean age of 3;9, and two elementary school groups (15 first graders with a mean age of 7;0, 22 third graders with a mean age of 9;3). In addition, within elementary school groups sequential and simultaneous bilinguals were compared. Pictorial stimuli were used to elicit narratives. These came from the MAIN, which allows for parallel elicitation of stories in both languages and is theoretically grounded in a story grammar approach that considers both the quantity of structural elements of the macrostructure produced and the quality (i.e., the production of the full episodes including goal, attempt, and outcome).

Developmental progression in macrostructure skills

For the development of story structure, story complexity, and ISTs, growth was predicted across the ages studied with greater growth between the preschool and first-grade groups in both languages. As expected, the findings showed more developmental growth in younger children in story structure and in story complexity for both languages. These results seem to be is in line with previous findings in monolingual development of narrative structure reported by Berman and Slobin

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

(1994) and Trabasso and Rodkin (1994) for children between 3 and 9 years of age. Both Berman and Slobin (1994) and Trabasso and Rodkin (1994) reported considerable development in children's production of story structure components and of their combination of components to produce more complex episodes. In addition, Kemper (1984), who reanalyzed three systems of evaluation of fantasy narratives in children between ages 2 and 10, found that by the age of 5, children already produce well-formed narratives, but narrative skill development goes on beyond this age. Although the monolingual children investigated in previous studies were younger than the children of the target group, for which the significant growth in the narrative skills is reported in this study, the considerable developmental progression in the narrative skills prior to age 7 seems to be similar. Although episodic structure is not yet (fully) acquired by age 5 (as has been shown for English narratives by Berman & Slobin, 1994; Kemper, 1984; Trabasso & Rodkin, 1994) and even by age 7 (as suggested by Pearson, 2002, and by the present study), its crucial components are already established in some children. The greater period of time separating preschoolers and first graders (39 months) versus the two school groups (27 months) might be an additional explanation of the more intensive growth of macrostructure observed between the preschoolers and the first graders. However, a calculation of the increase in the number of macrostructure components *per month* between the three groups still showed more intensive growth for both German and Russian between preschoolers and first graders (0.13 vs. 0.10 per month, respectively) than between first and third graders (German: 0.04; Russian: –0.02).

Thus, the study gives evidence that about half of first graders can produce at least one full episode, yet not necessarily the three GAOs that are required for a well-organized narrative in the MAIN. The ability to construct a full episodic structure suggests that there are some domains of narrative discourse that progress considerably between preschoolers and first graders. If one considers the developmental step from *no full episode* to *full episode* as a significant progression in narrative skills, this would suggest an even earlier, preschool, onset of acquisition than suggested by Karmiloff-Smith (1986) cited in Verhoeven (1993, p. 1): "discourse can be seen as the most significant domain of later language acquisition."

The strong ($r > .50$) and significant positive correlations of story structure but not ISTs between languages in the preschool and first graders ended by the third grade after children experienced two years of narrative teaching. This result suggests that the story structure is more invariant across languages in younger children (preschoolers and first graders) before explicit narrative teaching takes place, suggesting that cross-language associations of narrative skills might be sensitive to explicit teaching in one language. The similarity of story structure (prior to teaching) and of GAO production in both German and Russian in the preschoolers and first graders corroborates the findings of the few studies on bilinguals, namely, Pearson (2002) and Fiestas and Peña (2004) on Spanish–English bilinguals and Iluz-Cohen and Walters (2011) on English–Hebrew bilinguals and Gagarina et al. (2015) on further language pairs: story complexity might be invariant across dual languages.

In the GAO production, the highest level of story complexity, the dissociation across the languages of third graders, might be due to the effect of narrative teaching in L1. That is, although direct teaching how to combine the episodes'

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

single components into a complete GAO in L1 Russian might lead to its immediate and direct application in this language, the transfer to L2 might need some time. This is what the results seem to show: the skills of GAO production were invariant across languages prior to instruction, that is, in preschoolers and first graders. Then because of direct teaching in L1, story structure improved in the language of instruction (and thus a dissociation between L1 and L2 production of GAO took place), but had not yet been transferred to L2 (but note the difference from the general story complexity score). However, it is important to note that there was no significant improvement of GAO production between the first and third graders; thus, this interpretation is made cautiously. If GAO production skills are invariant, one can speculate that later on they should again be balanced across languages. The curriculum of the German and Russian lessons has been briefly compared: both curriculums of first graders start with learning how to read (letters, syllables, words, sentences, and texts). However, in Russian, special so-called *literaturnoje čtenie* "literary reading" lessons take place. Already in the first grade, the Russian textbooks introduce the notions of a text title and its connection to the main idea of the text and the plot and their constituents. Thus, explicit teaching of narrative episodic structure starts practically together with emerging literacy. This result supports previous findings on the direct and positive effects of narrative teaching for preschoolers (Hayward & Schneider, 2000) and suggests that an effect of teaching of GAO in one language may not be immediately transferred to the production of full episodes in another language, although narrative skills, as represented by the story complexity score, exhibit cross-language associations. Of course, a more thorough and systematic comparison of the curriculums should be done.

Although story length was not the main topic of the present article, a quick remark is necessary: the story length measured in a total number of word tokens improved significantly in both languages from preschoolers to first graders, but not from first graders to third graders. No direct comparisons of the narrative length between languages were performed, due to the differences in Russian with its absence of articles and a synthetic form for past tense as compared to German (cf. 12 words in German and only 5 words in Russian for the same sentence: A mother-bird flew to get the food for chicks, German: *Eine Vogelmutter ist weggeflogen um das Futter für die Küken zu holen*, Russian: *Mama uletela dobyvat' korm ptencam*).

The number of ISTs produced was almost identical in both languages in the preschool children with a maximum of four tokens produced in either language by any individual child. This picture changed by first grade (more IST tokens were produced in German than in Russian), and their number continued to increase in the oldest group. Because IST is a quantitative measure, and logically the number of ISTs produced is associated with the knowledge of respective words, I suggest that ISTs strongly depend on language (lexical) knowledge more than the other two measures of macrostructure. Thus, a more intensive increase in the ISTs in L2 German might be due to a more active development of L2 lexicon as compared to L1. L2 German is the societal language, which is shown to become more important for children with increasing age (Cobo-Lewis, Pearson, Eilers, & Umbel, 2002a, 2002b; Reich, 2007). Although the children in the present study are taught at

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

school in both languages (e.g., mathematics and music in German; history and physical education in Russian) and score within age-level expectations in both languages, German is the main language of communication between the children at school and in the media; it generally has more "weight" in everyday life. Thus a direct reflection of the societal dominance of German may be the number of ISTs produced by children with increasing age. This societal dominance did not appear to impact the use of overall story complexity, because story complexity can be said to be more language independent or invariant across language of a bilingual child.

This study showed that various constituents of narratives skills such as story structure, story complexity, and ISTs do not exhibit an identical developmental trajectory, suggesting the "fine-grained" composition of narrative skills. One of the possible future research directions might be the investigation of the associations of these different constituents and their role in the development of narrative skills. In addition, the more in-depth investigation of the microstructural elements (this study addressed only the length of stories in tokens) and their associations with the macrostructure should be performed.

Narrative skills of simultaneous and sequential bilinguals

In the second part of the study, some differences in the performance of sequential versus simultaneous elementary school children were found in favor of simultaneous bilinguals. The mean scores for sequential bilinguals were below those of simultaneous bilinguals on all three components of macrostructure, with this difference being significant in both languages for story complexity. While preliminary analyses showed that the performance of sequential versus simultaneous preschoolers (mean age of L2 onset = 30 months, mean length of exposure = 17 months) did not significantly differ on any measure, a gap seemed to develop during the next several years in the German narrative story structure (in Russian, there is a tendency to significance if one compares all simultaneous vs. sequential bilinguals; see Table 4) and in Russian and German story complexity of first- and third-grade children. However, an analysis of the highest complexity level only, GAO production, revealed significant differences only in Russian. For ISTs production, the most language-dependent component, because it evaluates the number of tokens, no differences were found between simultaneous and sequential bilinguals.

Generally speaking, sequential bilinguals in this study can be viewed as monolinguals for their first 2 years, that is, during the most crucial period for language acquisition (Grotjahn & Schlak, 2010), with the L2 coming later, after this period. It is age 2 that was seen by Piaget (1973) as the end of the sensorimotor and the beginning of preoperational stage in his seminal study on cognitive development in children. It is age of (L2 onset) 2, which had been proposed by Lenneberg (1967; cited by Singleton, 2005) to be an onset of a critical period that is characterized by a gradual decline in language acquisition capacity (due to the lateralization process): "children deafened before completion of the second year do not have any facilitation [in respect of oral skills] in comparison with the congenitally deaf," while those who had lost their hearing after that age "can be trained much more

easily in all the [oral] language arts" (Lenneberg 1967, p. 155). Singleton's (2005) review of an impact of age of L2 onset on L2 attainment emphasizes that the most frequently cited studies relate the changes in the acquisition capacity to the neurobiological changes in the brain and that there is no consensus on neurobiological representations of age of acquisition (cf. Muñoz & Singleton, 2011). This topic is beyond the scope of the present study and will not be treated further. However, the first 2 years separate simultaneous versus sequential bilinguals in the present study: for the sequential bilinguals, it is at this point that they transition from home to the nursery school, to a situation in which a child finds him/herself in a new environment and in a new peer group. With the entrance to a monolingual nursery school, the demands of learning a L2 come together with a new social situation that forces a child to communicate and to interact with a new set of caretakers and children. The constellation of these factors at age 2 might continue to influence story structure production in L2 German of sequential bilinguals in elementary school several years later. This effect is also seen in the overall evaluation of story complexity in both languages. Story complexity can be said to reflect a part of our cognitive abilities: the abilities to understand the intentions, build inferences, presuppose others' states of mind, and so forth. The finding that overall story complexity is somewhat lower in sequential bilinguals at school age might indirectly point to the fact that cognitive advantages are more pronounced in simultaneous bilinguals.

The presence of a significant difference in the overall story structure in both languages but no significant difference in GAO only in German across bilingual groups is somewhat surprising. However, it might be explained by the fact that the skill to combine goals, attempts, and outcomes is of a more "technical" nature than the ability to narrate the pictorial content, that is, to produce single components of the macrostructure, and is influenced by direct instruction. Because no direct instructions in L2 German were performed, the two types of bilinguals do not differ significantly in GAO production, whereas in L1 Russian, in which GAO teaching took place from the onset of literacy training, simultaneous bilinguals responded to the teaching more effectively having showed significantly better performance.

Educational and clinical implications

For story complexity, the evidence for significant improvement of this core narrative skill has implications for elementary school curriculum development. The results of this study, combined with knowledge of the direct impact of oral narrative skills on school and later academic success (Cameron, Hunt, & Linton, 1988) and improved performance as a result of narrative instruction (Hayward & Schneider, 2000), lead to the conclusion that targeted narrative instruction should be an important part of the primary school curriculum. In particular, narrative elicitation procedures used in MAIN are appropriate for these age groups and might be used in order to teach children how to tell narratives and to emphasize what constituents constitute a well-formed episode. This instruction probably should be performed in both languages. This is because direct teaching of narrative structure seemed to have a direct impact on narrative performance in the language in which it was taught (Russian in the present case), but did not undergo a quick cross-language

transfer to the other language, German in the case of this study. It would be interesting to investigate this contention in the future to find out how much time children need to transfer this taught knowledge from one language to another.

It is well established that "the early school years have been identified as an important time to evaluate children's prereadiness skills for reading acquisition" (Uccelli & Páez, 2007, p. 235). Professionals are encouraged to use narrative screening in the preschool and early school years and to treat vocabulary and other language skills as independent of narrative abilities, and to consider the association between two languages of a bilingual child when measuring narrative skills.

A further implication is motivated by the findings here for simultaneous versus sequential bilinguals: if both parents speak a language at home that differs from the language of the country where the family lives, children probably should be exposed to the school language as early as possible, prior to the age of 2 (cf. the first language first learning approach of Schwartz, 2013) with quantitatively and qualitatively rich input. The home language should be supported as well.

Because our bilingual children received education in both languages, the results may be compared to those of Collier (1989). The key finding of Collier's study is that simultaneous as well as sequential bilinguals perform better in school if they receive schooling in both languages. If bilinguals' primary language is not supported upon entering school, and children only receive schooling via their L2, they might be much less successful in school, especially as they move into the upper elementary and secondary years (for the synthesis of these results, see Collier & Thomas, 2009; Thomas & Collier, 2012).

Limitations of study

The small sample of participants in each age group limits the power of statistical inferences and the ability to draw generalizations. A further limitation concerns the MAIN instrument itself. Although MAIN and its measures (components of macrostructure) have been piloted with different children in various countries, it is still relatively untried and in need of additional testing. The suggested procedure for the evaluation of narrative skills can be said to be precise and effective only after much more data is analyzed.

Conclusion

In conclusion, despite the limitations of the study as far as the evaluation of bilingual narratives is concerned, for story structure, story complexity, and ISTs, we have been able to show more intensive development of narrative skills from preschoolers to first graders than from first graders to third graders and an advantage of simultaneous bilinguals in the production of story complexity. Many children as young as age 7 are able to construct a picture-based story that includes one full episode, including a goal, attempt, and outcome. Cross-language association of narrative skills, which was previously reported for other language pairs, was found for Russian–German bilinguals as well. However, not all three measures of macrostructure examined in this study showed significant correlations between L1 and L2. In particular, story structure, but not IST, can be said to be more

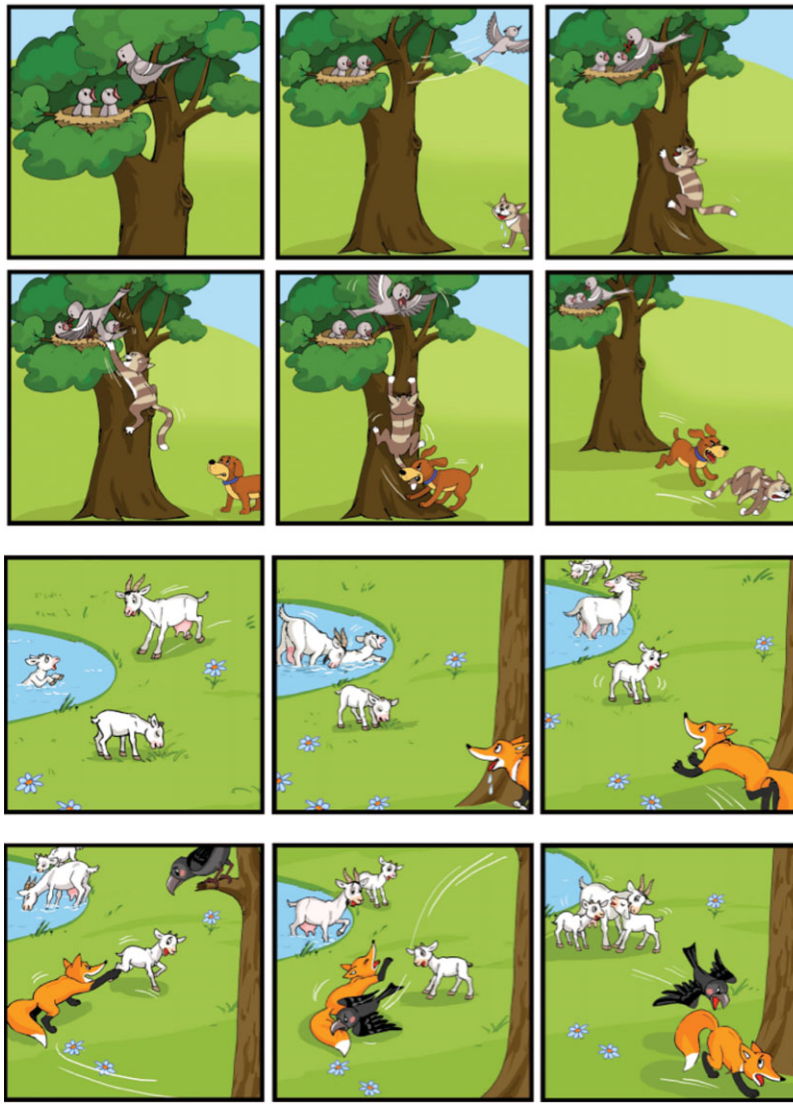
language invariant and is more likely to undergo language transfer in bilingual children. The last, but not the least, generalization shows that narrative complexity or organization is sensitive to teaching, and if it is taught in one language, it may not transfer immediately to the other language.

Productive narrative skills are not monolithic, but are composed of various aptitudes or abilities, such as story structure, story complexity, and ISTs. While in younger children they are not yet fully formed, the development of each component might follow a slightly different trajectory, depending on direct teaching, the degree of language dominance, and so on.

For future research, it is important to continue to examine the role of these three components of narrative skills and to investigate their degrees of language dependence. These three components offer an opportunity to more precisely identify invariance as well as variation across the two languages of a bilingual child and to compare simultaneous and sequential narratives' skills.

APPENDIX A

The Baby Birds (based on Hickmann, 2003) and Baby Goats (developed from Gülzow & Gagarina, 2007) are reduced-size copies of the picture stimuli. The original stimuli are in color (color online) and correspond to the A4 landscape format.



https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

The Cat and the Dog are reduced-size copies of the picture stimuli. The original stimuli are in color (color online) and correspond to the A4 landscape format.

https://doi.org/10.1017/S0142716415000430 Published online by Cambridge University Press

ACKNOWLEDGMENTS

The present study was supported in part by the German–Israeli Foundation (GIF) Grant 1113/2010, by Bundesministerium fuer Bildung und Forschung (BMBF) Grant 01UG0711, and by the Berliner Senate. I thank the anonymous reviewer for the helpful comments and suggestions and Elizabeth Kay Raining-Bird for her careful reading and many valuable suggestions on the working versions of this paper. I also thank all the children, their parents, and their teachers from the Berlin preschools and the primary school at the Brandenburg Gate for their participation in the study. For the organization of the complex synergy of testing for the preschool children and for complex statistics, my gratitude goes to the researchers of the Berlin Interdisciplinary Network for Multilingualism: Stefanie Düsterhöft, Dorothea Posse, Nathalie Topaj, and Felix Golcher. For the technical help I am deeply grateful to Natalie Sürmeli, Valerie Reichardt, and Marina Kubatina.

REFERENCES

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*, 291–317.

Bamberg, M. (1987). *The acquisition of narratives: Learning to use language*. Berlin: Mouton de Gruyter.

Benjamin, W. (Ed.). (1985). *Illuminationen*. Frankfurt am Main: Suhrkamp.

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

Bettge, S., & Oberwoehrmann, S. (2012). Migrationshintergrund als Einflussfaktor auf die kindliche Entwicklung im Einschulungsalter—Ein multivariates Modell. *Gesundheitswesen*, *74*, 203–209.

Bishop, D., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. *British Journal of Developmental Psychology*, *23*, 25–46.

Bishop, D. V. M. (1989). *Test for Reception of Grammar* (2nd ed.). Manchester: University of Manchester, Age and Cognitive Performance Research Centre.

Bishop, D. V. M., & Edmundson, A. (1987). Language impaired 4-year-olds: Transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*, 156–173.

Bliss, L. S., McCabe, A., & Miranda, E. A. (1998). Narrative assessment profile: Discourse analysis for school-age children. *Journal of Communication Disorders*, *31*, 347–363.

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

Bulheller, S., & Häcker, H. O. (Eds.). (2002). *Coloured Progressive Matrices (CPM): Deutsche Bearbeitung und Normierung nach J. C. Raven*. Frankfurt: Pearson Assessment.

Cameron, C., Hunt, A., & Linton, M. (1988). Medium effects on children's story rewriting and story retelling. *First Language*, *8*, 1–8.

Cobo-Lewis, A., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002a). Effects of bilingualism and bilingual education on oral and written Spanish skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 98–117). Clevedon: Multilingual Matters.

Cobo-Lewis, A., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002b). Effects of bilingualism and bilingual education on oral and written English skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 43–63). Clevedon: Multilingual Matters.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a school language. *TESOL Quarterly*, *23*, 509–531.

Collier, V. P., & Thomas, W. P. (2009). *Educating English learners for a transformed world*. Albuquerque, NM: Dual Language Education of New Mexico/Fuente Press.

Döpke, S. (2000). Introduction: On the status of cross-linguistic structures in research on young children growing up with two languages simultaneously. In S. Döpke (Ed.), *Cross-linguistic structures in simultaneous bilingualism* (pp. 1–10). Amsterdam: John Benjamins.

Fay, B. (1996). *Contemporary philosophy of social science: A multicultural approach*. Oxford: Wiley–Blackwell.

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

Fox, A. (2006). *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner Verlag GmbH.

Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., et al. (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. Silbereisen, P. Titzmann, & Y. Shavit (Eds.), *The challenges of diaspora migration: Interdisciplinary perspectives on Israel and Germany* (pp. 63–82). Farnham: Ashgate.

Gagarina, N., Klassert, A., & Topaj, N. (2010). Russian language proficiency test for multilingual children. *ZAS Papers in Linguistics* 54. Berlin: ZAS.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., et al. (2012). *MAIN: Multilingual Assessment Instrument for Narratives*. *ZAS papers in linguistics 56*. Berlin: ZAS.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., et al. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Language impairment testing in multilingual settings* (pp. 241–274). Amsterdam: Benjamins.

Gagarina, N., Voeikova, M., & Gruzincev, S. (2003). New version of morphological coding for the speech production of Russian children. In P. Kosta, J. Blaszczak, J. Frazek, L. Geist, & M. Zygis (Eds.), *Contributions of the Fourth European Conference on Formal Description of Slavic Languages* (pp. 243–258). Bern: Peter Lang.

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

Glasgow, C., & Cowley, J. (1994). *Renfrew Bus Story Test* (North American ed.). Centreville, DE: Centreville School.

Glueck, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige. WWT 6–10*. München: Elsevier/Urban & Fischer.

Grotjahn, R., & Schlak, T. (2010). Alter. In *Deutsch als Fremd- und Zweitsprache: Handbooks of linguistics and communication science* (Vol. 35, pp. 867–876). Berlin: de Gryuter.

Guelzow, I., & Gagarina, N. (2007). *Noun phrases, pronouns and anaphoric reference in young children narratives. ZAS papers in linguistics 48* (pp. 203–223). Berlin: ZAS.

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, *13*, 175–197.

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics*, *29*, 3–20.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. New York: Basic Books.

Hamann, C., Chilla, S., Ruigendijk, E., & Abed Ibrahim, L. (2013). *A German sentence repetition task: Testing bilingual Russian-German children*. Poster presented at the COST Action IS0806 Conference, Crakow.

Hayward, D., & Schneider, P. (2000). Effectiveness of teaching story grammar knowledge to preschool children with language impairment: An exploratory study. *Child Language Teaching and Therapy*, *16*, 255–284.

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech–Language Pathology*, *19*, 154–166.

Hickmann, M. (2003).*Children's discourse: Person, space and time across languages*. Cambridge: Cambridge University Press.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*, 58–74.

Jia, R., Yiu, E., Sorenson Duncan, T., & Paradis, J. (2010). *Story grammar in the narratives of Chinese heritage language children*. Talk presented at the First International Conference on Heritage/Community Language, Los Angeles.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistics and repair data. *Cognition*, *23*, 95–147.

Kemper, S. (1984). The development of narrative skills: Explanations and entertainments. In S. A. Kuczaj (Ed.), *Discourse development* (pp. 99–124). New York: Academic Press.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Li, P. (2013). Successive language acquisition. In F. Grosjean & P. Li (Eds.), *Psycholinguistics of bilingualism* (pp. 201–210). Oxford: Wiley–Blackwell.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research*, *36*, 868–882.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, *38*, 415–425.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.

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

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

McCabe, A. (1996). Evaluating narrative discourse skills. In K. Cole, P. Dale, & D. Thal (Eds.), *Assessment of communication and language* (pp. 121–141). Baltimore, MD: Paul H. Brooks.

McCabe, A., & Peterson, C. (1984). What makes a good story? *Journal of Psycholinguistic Research*, *13*, 457–480.

McCabe, A., & Rollins, P. R. (1994). Assessment of preschool narrative skills. *American Journal of Speech–Language Pathology*, *3*, 45–56.

Meir, N., Armon-Lotem, S., & Walters, J. (2015). *Disentangling bilingualism from SLI using sentence repetition tasks: The impact of L1 and L2 properties.* Manuscript submitted for publication.

Meisel, J. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*, 5–34.

Mohades, S. G., Struys, E., Van Schuerbeek, P., Mondt, K., Van De Craen, P., & Luypaert, R. (2012). DTI reveals structural differences in white matter tracts between bilingual and monolingual children. *Brain Research*, *1435*, 72–80.

Montanari, S. (2004). The development of narrative competence in the L1 and L2 of Spanish-English bilingual children. *International Journal of Bilingualism*, *8*, 449–497.

Montrul, S. A. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, *44*, 1–35. doi:10.1017/S026144481000032

Nash, C. (1994). *Narrative in culture: The uses of storytelling in the sciences, philosophy and literature*. London: Routledge.

Newton, A. (1997).*Narrative ethics*. Cambridge, MA: Harvard University Press.
Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, *36*, 125–138.
Norris, J. A., & Bruning, R. H. (1988). Cohesion in the narratives of good and poor readers. *Journal of Speech and Hearing Disorders*, *53*, 416–424.
Paradis, J. (2008). *Are simultaneous and early sequential bilingual acquisition fundamentally different?* Paper presented at Models of Interaction in Bilinguals Conference, University of Wales.
Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics*, *31*, 227–252.
Paradis, J., Genesee, F., & Crago, M. B. (2010). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Baltimore, MD: Paul H. Brookes.
Pearson, B. Z. (2002). Language and mind in the stories of bilingual children. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (pp. 135–174). Amsterdam: John Benjamins.
Peterson, C., & McCabe, A. (1991). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum Press.
Piaget, J. (1973). *Main trends in psychology*. London: George Allen & Unwin.
Polkinghorne, D. E. (1988). *Narrative knowing and the human sciences*. Albany, NY: State University of New York Press.
Reich, H. (2007). Forschungsstand und Desideratenaufweis zu Migrationslinguistik und Migrationspädagogik für die Zwecke des "Anforderungsrahmens." In Bundesministerium für Bildung, Forschung, & K. Ehlich (Eds.), *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund* (pp. 121–170). Berlin: BMBF.
Reich, H., & Roth, H.-J. (2004). *Hamburger Verfahren zur Analyse des Sprachstands Fünfjähriger*. HAVAS 5. Hamburg: Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg.
Ringmann, S., & Siegmüller, J. (2013). Die Beziehung zwischen Satzgrammatik und Erzählfähigkeit im unauffälligen und auffälligen Spracherwerb. *Forschung Sprache*, *1*, 36–50.
Ruberg, T. (2013). Problembereiche im kindlichen Zweitspracherwerb. *Sprache. Stimme. Gehör*, *37*, 181–185.
Schneider, P., Dubé, R. V., & Hayward, D. (2005). *The Edmonton Narrative Norms Instrument.* Retrieved from http://www.rehabresearch.ualberta.ca/enni
Schneider, P., Hayward, D., & Dube, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech Pathology and Audiology*, *30*, 224–238.
Schulz, P., & Tracy, R. (2011). *Linguistische Sprachstandserhebung—Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe.
Schwartz, M. (2013). The impact of the first language first model on vocabulary development among preschool bilingual children. *Reading and Writing*, *27*, 709–732.
Simon-Cereijido, G., & Gutiérrez-Clellen, V. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, *30*, 315–337.
Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, *43*, 269–285.
Skerra, A., Adani, F., & Gagarina, N. (2013). Diskurskohäsive Mittel in Erzählungen als diagnostischer Marker für Sprachentwicklungsstörungen. In T. Fritzsche, C. B. Meyer, A. Adelt, & J. Roß (Eds.), *Spektrum patholinguistik* (Vol. 6, pp. 127–158). Potsdam: Universitätsverlag Potsdam.
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). Norwood, NJ: Ablex.

Swanson, L., Fey, M., Mills, C., & Hood, L. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech–Language Pathology*, *14*, 131–143.

Tellegen, P. J., Laros, J. A., & Petermann, F. (2007). *Snijders–Oomen Non-verbaler Intelligenztest von 2½ bis 7 Jahren (SON-R 2½–7)*. Göttingen: Hogrefe.

Thomas, W. P., & Collier, V. P. (2012). *Dual language education for a transformed world*. Albuquerque, NM: Dual Language Education of New Mexico/Fuente Press.

Torrance, N., & Olson, D. R. (1984). Oral language competence and the acquisition of literacy. In A. Pellegrini & T. Yawkey (Eds.), *The development of oral and written language in social contexts* (pp. 167–181). Norwood, NJ: Ablex.

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

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goals/plans: A conceptual basis for narrating *Frog, where are you?* In R. Berman & D. Slobin (Eds.), *Relating events in narrative: A cross-linguistic developmental study* (pp. 85–106). Hillsdale, NJ: Erlbaum.

Uccelli, P., & Páez, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*, 225–236.

van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In P. van den Broek, P. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension: Bridging fictional and actual events* (pp. 321–342). Mahwah, NJ: Erlbaum.

Verhoeven, L. (1993). Acquisition of narrative skills in a bilingual context. In B. Kettemann & W. Wieden (Eds.), *Current issues in European second language acquisition research* (pp. 307–323). Tuebingen: Gunter Narr.

Wallach, G. P. (2008). *Language intervention for school-age students: Setting goals for academic success*. St. Louis, MO: Mosby.

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