Macrostructure in the narratives of Indonesian-Dutch bilinguals

Relation to age and exposure

Elena Tribushinina, Mila Irmawati, and Pim Mak
Universiteit Utrecht

There is no agreement regarding the relationship between narrative abilities in the two languages of a bilingual child. In this paper, we test the hypothesis that such cross-language relationships depend on age and language exposure by studying the narrative skills of 32 Indonesian-Dutch bilinguals (mean age: 8;5, range: 5;0–11;9). The narratives were elicited by means of the Multilingual Assessment Instrument for Narratives (MAIN) and analysed for story structure, episodic complexity and use of internal state terms (ISTs) in the home language (Indonesian) and majority language (Dutch). The results demonstrate that story structure scores in the home language (but not in the majority language) were positively related to age. Exposure measures (current Dutch/Indonesian input, current richness of Dutch/Indonesian input, and length of exposure to Dutch) did not predict the macrostructure scores. There was a significant positive cross-language relationship in story structure and episodic complexity, and this relationship became stronger as a function of length of exposure to Dutch. There was also a positive cross-lingual relation in IST use, but it became weaker with age. The results support the idea that narrative skills are transferable between languages and suggest that cross-language relationships may interact with age and exposure factors in differential ways.

Keywords: story grammar, structural complexity, internal state terms, bilingualism, cross-language transfer

1. Introduction

The development of narrative ability is a crucial aspect of child (language) development. Comprehension of wordless picture narratives and production of stories based on a picture series are good predictors of developing literacy skills (Bishop & Edmundson, 1987; Paris & Paris, 2003). More recently, narrative assessments have

https://doi.org/10.1075/lab.20015.tri | Published online: 16 February 2021
Linguistic Approaches to Bilingualism 12:4 (2022), pp. 540–570. ISSN 1879-9264 | E-ISSN 1879-9272

also proven useful for disentangling the effects of reduced language exposure (due to bilingualism) and reduced uptake (due to language disorders) (e.g., Altman, Armon-Lotem, Fichman & Walters, 2016; Boerma, Leseman, Timmermeister, Wijnen & Blom, 2016; Cleave, Girolametto, Chen & Johnson, 2010; Fichman, Altman, Voloskovich & Armon-Lotem, 2017; Govindarajan & Paradis, 2019; Squires, Lugo-Neris, Peña, Bedore, Bohman & Gillam, 2014).

A general finding is that narratives produced in the weaker (non-dominant) language of a bilingual child may be relatively weak at the level of *microstructure*, i.e. in the use of lexical and grammatical elements to produce a coherent story. The weaker language can be either a majority language (ML) spoken in the country of residence, or a home language (HL) that the child primarily uses with family members. Simultaneous bilinguals acquiring both languages from birth (De Houwer, 2009) are usually dominant in the ML, because they receive a lot of input in the societal language, whereas their exposure to the other language (not maintained by the society) tends to be significantly reduced, especially when schooling in the ML starts (Cleave et al., 2010; Tribushinina, Mak, Andreiushina, Dubinkina & Sanders, 2017). In child second language (L2) learners, whose first exposure to the ML is around age 3 (McLaughlin, 1978) or 4 (Unsworth, 2005), the ML is often the weaker language due to a later onset and reduced cumulative exposure (Unsworth, 2013). However, with age and longer exposure, the L2 may gradually take over and become the dominant language of a child (Montrul, 2010). Hence, microstructure scores depend on the child's dominance patterns and proficiency in the two languages.

In contrast to microstructure, narrative *macrostructure* – general organisation and episodic complexity of a story – appears less sensitive to reduced language exposure, because it hinges on more general cognitive and pragmatic skills, such as understanding goal-directed behaviour of the characters. Macrostructure is usually operationalized as the number of story grammar elements (e.g., setting, initiating event, internal response, goal(s), attempt(s), consequence(s) and resolution) and complexity of episodes (e.g., complex goal-attempt-outcome sequences *vs.* incomplete attempt-outcome or goal-attempt sequences). Some analyses of macrostructure also include internal state terms (ISTs) (e.g. Altman et al., 2016; Boerma et al., 2016; Kunnari, Välimaa & Laukkanen-Nevala, 2016). A bilingual child who knows how to structure a coherent narrative in her stronger language, be it the first language (L1) of sequential bilinguals or the societal language of simultaneous bilinguals, may use this knowledge in producing narratives in the other language as well. This cross-language relationship is commonly conceptualized as positive transfer (Cummins, 1979; Govindarajan & Paradis, 2019; Pearson, 2002; Petersen, Thompsen, Guiserson & Spencer, 2016; Squires et al., 2014).

Even though there seems to be general agreement that narrative skills can be transferred between languages, empirical evidence regarding such cross-language relationships in narrative abilities is inconclusive. Several studies report similar macrostructure scores in the two languages of bilingual children (Altman et al., 2016; Boerma et al., 2016; Bohnacker, 2016; Cleave et al., 2010; Fichman et al., 2017; Fiestas & Peña, 2004; Govindarajan & Paradis, 2019; Lindgren, 2018). For example, Fiestas and Peña (2004) found equal narrative complexity in L1 Spanish and L2 English, which they interpret as interrelated narrative skills. Likewise, Fusté-Herrmann, Silliman, Bahr, Fasnacht and Federico (2006) report similar rates of mental state terms in the narratives of Spanish monolinguals and Spanish-English bilinguals. These studies seem to suggest that the understanding of protagonists' goal-directed behaviour and the ability to structure a narrative in a coherent manner are general cognitive skills that should be invariant and transferable across languages.

But there are also studies revealing poorer performance in one of the languages, suggesting proficiency-related limitations (Fichman et al., 2017; Kapalková, Polišenská, Marková & Fenton, 2016; Gutiérrez-Clellen, 2002; Montanari, 2004). Hence, a child with a limited proficiency in one of the languages may well be able to produce a coherent account of goal-directed sequences in her stronger language, but may not have enough resources for producing a similar narrative in the weaker language, even if s/he is equipped with the relevant metalinguistic/cognitive skills. For example, Kapalková and colleagues (2016) studied narrative macrostructure in L1 and L2 stories produced by Slovak pre-schoolers exposed to L2 English through immersion nurseries. The results revealed higher macrostructure scores in L1 Slovak than in L2 English. In a similar vein, Montanari (2004) traced the development of narrative competence in three Spanish-English bilinguals with varying levels of L2 proficiency and concluded that insufficient knowledge of "formal means in the children's weaker language prevented them from constructing narratives with an explicitly motivated beginning (a problem), an elaboration or complicated reaction of the hero to this problem, and a final resolution" (p. 494). In contrast, Uccelli and Paéz (2007) report better narrative performance in L2 English than in L1 Spanish, but the quality of L1 narratives at kindergarten was predictive of the quality of L2 narratives in the first grade, suggesting possible transferability of L1 narrative skills.

It is plausible to assume that the controversial results might be attributed to a number of differences between the studies. For one, the existing studies differ in the tasks that have been employed. Some used story retells (e.g., Altman et al., 2016; Boerma et al., 2016; Squires et al., 2014) or video retells (Silliman, Bahr, Brea, Hnath-Chisolm & Mahecha, 2002), whereas others employed story telling tasks (e.g., Govindarajan & Paradis, 2019; Gutiérrez-Clellen, 2002), or a combination

of telling and retelling (e.g., Gagarina, 2016; Kunnari et al., 2016). Different elicitation modes may have produced different results. Gutiérrez-Clellen (2002) compared narrative recall and narrative production of bilingual children (aged 7–8) in L1 Spanish and L2 English. The results revealed that children who had difficulty with narrative recall in L2 English, were still able to produce good English narratives in a story generation task. The author suggests that this discrepancy might be due to difficulties with processing L2 word order in the recall task. In contrast, Kunnari and colleagues (2016) found no differences between Finnish monolinguals and Finnish-Swedish bilinguals in story structure when a retell task was used, but monolinguals outperformed the bilinguals on the telling task, presumably due to greater cognitive demands involved.

Another essential difference is that different studies included different components of narrative ability. Importantly, there is evidence suggesting that narrative skills are complex and the development of different subskills may reveal differential trajectories. For example, Gagarina (2016) found significant cross-language correlations for story structure, but not for story complexity and IST use. Relatedly, Kunnari et al. (2016) report no differences between bilinguals and monolinguals in story complexity and IST use, but monolinguals outperformed bilinguals on story structure (when measured by a telling task).

Finally, prior studies tested bilinguals of different ages and different types of bilinguals. The ages of the participants in the existing studies of narrative macrostructure varied from pre-schoolers (e.g., Kapalková et al., 2016) to school-aged children (e.g., Gutiérrez-Clellen, 2002) and adults (e.g., Krasnoshchekova & Kashleva, 2019). Some studies looked at simultaneous bilinguals (e.g., Bohnacker, 2016; Kunnari et al., 2016), whereas other studies focussed on child L2 learners (Altman et al., 2016; Gutiérrez-Clellen, 2002; Kapalková et al., 2016; Petersen et al., 2016; Schwartz & Shaul, 2013; Squires et al., 2014) or included a combination of simultaneous and sequential bilinguals (Gagarina, 2016; Lindgren, 2018; Öztekin, 2019). Importantly, the relationship between narrative skills in the two languages of a bilingual is likely to change as a function of age and language exposure. For instance, if a child only receives formal schooling in the ML, she may simply not have enough linguistic resources to talk about certain kinds of scripts in the HL (Schwartz & Shaul, 2013) or produce as many ISTs as in the dominant ML (Altman et al., 2016). This paper further pursues this avenue in order to explore whether the (strength of the) relationship between narrative skills in HL and ML depends on the child's age and language exposure. Before reporting the experiment, we briefly review the currently available (mainly indirect) evidence that these variables may mediate the relationship between narrative competence in the two languages of a bilingual.

2. Factors influencing cross-language relationships

2.1 Biological age

The cross-lingual relationships in the domain of narrative macrostructure may be contingent on the child's biological age, as older learners have more cognitive and (meta)linguistic resources available for transfer. More specifically, children develop knowledge of basic story grammar elements with age (Kunnari et al., 2016; Paris & Paris, 2003). Furthermore, the ability to produce a good narrative entails that a child not only describes a sequence of events, but also creates a coherent account of the protagonists' thoughts, motives and internal states. In Curenton's words, "in order to be a good story-teller and story-listener children must be capable of simultaneously understanding not only *what* has happened in the story but also *why* it has happened" (Curenton, 2011, p. 791). Hence, the development of narrative ability is also contingent on the emerging Theory of Mind and perspective-taking skills. Accordingly, age-related improvements in the use of mental state terms and a positive relationship between a child's cognitive status and amount of internal state talk have been repeatedly reported in the literature (Chen & Yan, 2011; Curenton, 2004, 2011; Curenton & Justice, 2004; Ukrainetz, Justice, Kaderavek, Eisenberg, Gillam & Harm, 2005).

In summary, in older children, a whole array of linguistic, pragmatic and cognitive skills is more developed, and therefore there is more basis for positive cross-linguistic transfer. The present study will shed more light on this issue by studying whether (and how) age influences the relationship between story structure, story complexity and IST use in the two languages of dual language learners.

2.2 Language exposure

For children whose ML is not their L1, biological age may be a less important predictor of narrative quality than exposure to ML. As explained above, proficiency-related limitations may impede transferability of narrative skills between languages (Fichman et al., 2017; Gutiérrez-Clellen, 2002; Kapalková et al., 2016; Montanari, 2004). Take, for example, a 5-year-old who has had reduced exposure to the ML because she only started acquiring it at age 4. This child may well be able to produce a good narrative in her HL, but she will probably not have enough linguistic resources available to her in the L2 to tell a story of comparable structure and complexity. In other words, the transferability of narrative skills is likely to be impeded by limited proficiency in the weaker language. Hence it may be predicted that the amount of positive transfer of narrative skills will be positively related to the length, amount and richness of L2 exposure.

Thus far, the possible role of language exposure in the cross-language relationships in narrative competence has not been explored in the literature. This said, several studies have looked at exposure factors as predictors of narrative competence *within* languages. However, their findings are rather contradictory. For instance, the results reported by Schwartz and Shaul (2013) have revealed that later L2 exposure does not impede the development of script schema knowledge in L2. In contrast, Gagarina (2016) has found higher story structure scores in German for simultaneous bilinguals compared to early L2 learners of German, suggesting that shorter exposure might have put early sequential bilinguals at a disadvantage. Similarly, Govindarajan and Paradis (2019) have demonstrated that length of L2 exposure and richness of exposure predicted story grammar scores in typically-developing L2 learners of English. At the same time, they found no positive relationship between the amount of L2 use at home and narrative outcomes. In the same vein, Hipfner-Boucher, Milburn, Weitzman, Greenberg, Pelletier and Girolametto (2015) found no differences in macrostructure scores between English L2 learners who mainly spoke L2 English at home and those who mainly spoke their L1 at home. These results are in contrast to the findings by Squires and colleagues (2014) demonstrating that the amount of L2 exposure was positively related to macrostructure scores in the L2 (English), but not in the L1 (Spanish).

In summary, it is not yet clear how length and amount of exposure are related to narrative outcomes in ML and HL. The present study will contribute to this line of research. Additionally, we will explore whether (and how) length of exposure to ML may influence the relationship between ML and HL narrative skills.

3. The present study

This research focuses on the macrostructure of narratives produced by Indonesian-Dutch bilinguals residing in the Netherlands and acquiring Dutch as a ML and Indonesian as a HL. All children were exposed to Indonesian from birth and Dutch from a later age (about half before age 3 and half after age 3).

The first goal of this research is to establish whether macrostructure scores in HL and ML are predicted by biological age and input factors (length of exposure to Dutch, current Indonesian/Dutch input and current richness of Indonesian/Dutch input). As discussed above, there is a lot of evidence that narrative competence increases with age (e.g., Chen & Yan, 2011; Curenton, 2004, 2011; Gagarina, 2016; Ukrainetz et al., 2005). Hence, we predict a positive relationship between age and narrative scores in the HL (which was L1 for all the participants) across all three macrostructure measures. The relationship between age and ML performance is likely to be less straightforward, because age of Dutch onset

varied widely across the sample, and because amount and length of L2 exposure have been shown to outweigh the effects of biological age in L2 acquisition (Thordardottir, 2019). Regarding the relationship between exposure and narrative competence, evidence in the literature is highly controversial. Length, amount and richness of exposure have all been found to contribute to the development of narrative skills in some studies, but not others (see Section 2.2). Therefore, a null hypothesis will be adopted.

Secondly, this study aims to determine whether ML (Dutch) macrostructure scores can be predicted from the corresponding scores in the HL (Indonesian). Based on prior research, we hypothesize that the macrostructure scores in the two languages are related, in the sense that good story-tellers in the HL are also good story-tellers in the ML (Gagarina, 2016; Lindgren, 2018; Öztekin, 2019).

Finally, this study aims to shed more light on whether (and in what ways) the relationship between narrative skills in the two languages may interact with biological age and length of ML exposure. So far, such explorations have been scarce (e.g., Govindarajan & Paradis, 2019). Based on the studies showing that beginning L2 learners may not have enough linguistic resources to benefit from transfer of L1 macrostructure (e.g., Kapalková et al., 2002; Montanari, 2004), we predict that the relationships between narrative skills in Indonesian and Dutch may become stronger with increasing exposure to Dutch. We also expect stronger HL-ML relationships in older children, since they possess more developed linguistic, pragmatic and cognitive skills and, therefore, more basis for positive cross-lingual transfer.

4. Method

This research was screened and approved by the Ethical Assessment Committee of the Utrecht Institute of Linguistics. The parents of all participants signed informed consent.

4.1 Participants

Forty-four bilingual children (18 female) were recruited for this research. They all had normal hearing and no history of language disorders. Twelve participants were excluded from the analyses because they refused to tell a story in Dutch (3 three-year-olds, 3 four-year-olds, 4 five-year-olds, 1 six-year-old and 1 nine-year-old). These participants had to be excluded because it is not possible to study cross-language relationships if only one of the languages is present in the dataset. The final dataset included 32 participants: 5 five-year-olds, 5 six-year-olds, 3 seven-

year-olds, 3 eight-year-olds, 6 nine-year-olds, 6 ten-year-olds and 4 eleven-year-olds. Table 1 presents participant characteristics. This information was gathered by means of a parental questionnaire about children's language histories and language use in the family (see Section 4.2.1).

Twenty of the participants were born in Indonesia and 12 were born in the Netherlands. Two participants attended an international school, the remaining 30 children attended a mainstream primary school in the Netherlands. The participants came from mid- to high SES families and lived in cities and towns. All fathers and 26 mothers were university graduates. All mothers and 30 fathers were Indonesian (one father was Dutch and one was Vietnamese). All Indonesian parents were first-generation immigrants. One family came from South Sulawesi and one child's mother came from Lombok; all other parents were from Java. All children were exposed to colloquial Indonesian at home. Additionally, seven children also heard the Javanese dialect, one child heard Sundanese and one participant was exposed to both Javanese and Sundanese.

Table 1. Participant characteristics[1]

Characteristic	Mean	SD	Range
Age at testing (months)	101.8	24.8	60–141
Age of onset Dutch (months)	46.2	30.9	5–110
Length of exposure to Dutch (months)	55.5	39.7	5–135
Exposure to Indonesian before age 4 (max = 4)	3.8	1.9	3–4
Exposure to Dutch before age 4 (max = 4)	1.9	1.8	0–4
Current Indonesian input (max = 28)	17.8	5.7	7–26
Current Dutch input (max = 28)	16.2	4.9	6–26
Current richness of Indonesian input (max = 9)	6.0	1.5	3–8
Current richness of Dutch input (max = 9)	7.1	1.2	4–9

Seventeen participants had their first exposure to Dutch before age 3 (5 of them below age 1) and 15 participants came into contact with Dutch after age 3 (4 of them after age 7). Length of exposure to Dutch was distributed as follows: under 2 years ($N = 8$), 2–5 years ($N = 13$), 6–11 years ($N = 11$).

Before the age of 4 years, the participants had more exposure to Indonesian than to Dutch: $t(31) = 5.53$, $p < .001$. However, the amounts of current exposure

1. The input measures represent composite scores derived from the parental questionnaire. The scoring method is described in Section 4.2.1.

to ML and HL did not differ significantly: $t(31) = 0.91$, $p = .37$. Ten children were reported to have less than 45% of exposure to Indonesian (and more than 55% to Dutch), 6 children had balanced exposure to both languages (around 50% each), and 16 children were exposed to Indonesian more than 55% of the time.[2] According to parental reports, 21 children felt more comfortable using Indonesian than Dutch, 8 children were reported to feel more at home in Dutch and 3 participants felt equally comfortable in both their languages.

4.2 Instruments and procedures

This study used two instruments developed within the framework of the COST Action IS0804 *Language Assessment in a Multilingual Society: Linguistic Patterns and the Road to Assessment.*

4.2.1 *Questionnaire for parents of bilingual children (PaBIQ)*

An Indonesian version of the Questionnaire for Parents of Bilingual Children (PaBiQ, Tuller, 2015) was created to gather detailed information about the children's language development and language use. The parents filled in the questionnaire while the child participated in the narrative tasks.

Three measures derived from the questionnaire were included in the analyses (see Table 1 for mean scores and ranges). *Length of ML exposure* was calculated by subtracting the age of first exposure to Dutch (as determined by the question *At what age was your child first in contact with each of his/her languages?*) from age at testing.

Current Indonesian/Dutch input was a measure of amount of current exposure to each of the languages (Blom, Boerma, Bosma, Cornips, Van den Heuij & Timmermeister, 2020),[3] operationalized here as a composite score for the questions regarding input from the mother, the father, another adult who regularly took care of the child, siblings, child's friends and family friends (0 = never; 1 = rarely; 2 = sometimes; 3 = usually; 4 = very often/always) and the scores for weekly activities, including reading, watching television/movies and storytelling (0 = never or almost never; 1 = at least once a week; 2 = every day). The maximum possible score was 28. For example, participant 001 very often used Indonesian

2. These percentages were calculated relative to the total composite exposure score for both languages (see Section 4.2.1).

3. We follow Blom et al. (2020) in the use of this term. Other terms used for the same or similar construct in the literature include *intensity of exposure* (Blom & Bosma, 2016), *language use in the home* (Fichman et al., 2017; Paradis, Tulpar & Arppe, 2016) and *current exposure* (Boerma et al., 2016).

with the mother (4), father (4), another adult taking care of the child (4), siblings (4), friends (4) and family friends (4). She never used Indonesian for reading (0) and storytelling (0), but sometimes watched TV programmes in Indonesian (1). Hence, this child's composite exposure score was 25.

Current richness of Indonesian/ Dutch input was operationalized as the number of speaker types (mother, father, another adult, siblings, friends, family friends) and the number of weekly activities (reading, film/video watching, storytelling) for each language, irrespective of frequency. The maximum score for richness was 9. In the above example (participant 001), the richness score in Indonesian was 7.

4.2.2 *Multilingual assessment instrument for narratives (MAIN)*

Narratives were elicited via the MAIN story-telling task (Gagarina, Klop, Kunnari, Tantele, Valimaa, Balciuniene, Bohnacker, Walters, 2012). MAIN comprises four sets of six coloured pictures (Cat, Dog, Baby Birds and Baby Goats), representing four stories with identical story and episodic structure. All narratives contain three episodes. For example, in episode 1 of the Cat/Dog Story, a cat/dog wants to catch a butterfly/mouse (goal), leaps/runs to catch it (attempt), but falls into the bush/bumps into the tree (outcome). In the meantime, a boy comes; he carries a ball/balloon in one hand and fish/sausages in the other hand. In episode 2, the boy loses his ball/balloon, wants to get it back (goal), tries to pull the ball/balloon out of the lake/tree (attempt) and finally gets his ball/balloon back (outcome). Meanwhile, the cat/dog notices the fish/sausages, wants to steal them (goal), attempts to do so (attempt) and eats one fish/sausage (outcome).

According to the MAIN protocol, MAIN1 (Cat/Dog) should be used for retells and MAIN2 (Baby Birds/Goats) for a story generation (telling) task. However, we decided to use a telling task only, since our participants were quite old and story generation appears to be more demanding than retells (Kunnari et al., 2016). The Cat Story and the Dog Story were elicited in Indonesian. The stories about Baby Birds and Baby Goats were elicited in Dutch. MAIN1 and MAIN2 have been found to have equal levels of difficulty and reveal no task effects when used for story generation: Both monolingual and bilingual children have been shown to attain similar macrostructure scores on MAIN1 and MAIN2 (Lindgren, 2018; Öztekin, 2019). We included Story in the random part of our statistical models (see below) in order to control for possible task differences.

Since there is no Indonesian version of MAIN, we used the Dutch version as a basis for the translation into Indonesian. The order of the narratives was counterbalanced among participants. An anonymous reviewer raised a concern about the adequacy of using picture books for eliciting narratives in Indonesian, since using picture books for storytelling is less common in Indonesia (Klamer & Moro,

2020). Notice, however, that our participants were raised in the Netherlands and attended regular Dutch schools, where they had sufficient experience with books. Also, parents reported that their children regularly engaged in book-reading and storytelling in both languages (see Section 6.3).

All participants were tested in their homes in two sessions. Each session took 10–15 minutes. Unfortunately, for practical reasons, it was not possible to have an interval of seven days between the test sessions in the two languages. Both sessions took place on the same day and were conducted by the second author of this paper. According to the MAIN protocol, the order of languages should be counter-balanced. However, we decided to elicit narratives first in Indonesian and then in Dutch. This is because this order appeared more natural given the fact that all children were tested at home and the language used in the homes was predominantly Indonesian. From the moment the researcher entered the home until the onset of testing, the overall language mode was Indonesian and we decided not to break this mode by introducing the first task in a different language. Furthermore, as reported in Section 4.1, the majority of the participants felt more at home using Indonesian. After producing two Indonesian narratives, the participants felt more comfortable performing a similar task in their weaker language. After session 1, they were told that after a break they would perform a similar task in Dutch.

The narratives were elicited following the general procedures specified in the MAIN guidelines (Gagarina et al., 2012). The experimenter sat opposite to the child. The participant was given several envelopes containing one of the four picture strips and was asked to choose one envelope. This was done in order avoid the assumption of shared knowledge. Then the child was asked to take out the pictures, unfold the strip and look through the story. After that, the participants were asked to fold the strip again and to start telling the story without showing the pictures to the investigator. A standard fold-out procedure was followed: The participant started telling a story based on the first two pictures, then unfolded the next two pictures and finally the last two pictures. This procedure is less taxing for working memory and allows children to cover all three episodes. The investigator intervened as little as possible. If the child fell silent, neutral encouraging prompts (e.g., *Ga door* 'Go on'; *Nog iets?* 'Anything else?') were used.

The narratives were audio-recorded and later transcribed in a CHAT format using the CLAN (Computerized Language Analysis) software (MacWhinney, 2000). The Indonesian stories were transcribed and scored by the second author, in regular consultation with the first author (and sometimes with the creators of MAIN, p.c.). The Dutch narratives were transcribed and scored by a trained research assistant, in regular consultations with all the authors of this paper. All cases of doubt and disagreement were thoroughly discussed by the team and resolved by consensus.

The transcriptions were analysed for story structure, structural complexity and IST use following the MAIN scoring protocol (Gagarina et al., 2012).

Story structure captures how many story grammar elements were included in the narrative. Following the MAIN scoring protocol, the story elements scored were setting (time and place) and, for each of the three episodes, IST as initiating event (IST-IE), goal (G), attempt (A), outcome (O), and IST as a response (IST-R). The maximum score for story structure was 17 (2 for setting and 5 for each of the three episodes). Sample narratives tagged for story structure are provided in the Appendix.

Structural complexity is a measure of well-formedness of episodes. A complete episode consists of a goal, an attempt to reach the goal and an outcome (GAO). For each narrative we counted the number of complete episodes (GAO), sequences (AO/GO/GA) and single components (A/O/G). In order to calculate a complexity score, 3 points were given to each complete episode (GAO) and 2 points were awarded for each sequence consisting of two elements (AO/GO/GA). Single attempts and outcomes did not receive any points. However, we decided to award 1 point for each separate G, since goals are more complex than attempts and outcomes, and are mastered relatively late (Bohnacker, 2016; Trabasso & Nickels, 1992). The maximum possible number of points for structural complexity was 9 per narrative.

Finally, the total number of tokens of ISTs per narrative was counted. ISTs included perceptual state terms (e.g., *melihat* 'see'), physiological state terms (e.g., *lapar* 'hungry'), consciousness terms (e.g., *nyadar* 'realize'), emotion terms (e.g., *senang* 'happy'), mental verbs (e.g., *ingin* 'want') and verbs of saying (e.g., *berkata* 'say').

4.3 Data analysis

The data were subjected to linear regression analyses, using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). For both Indonesian and Dutch narratives, we first investigated whether the scores on the three dependent measures (story structure, structural complexity and IST frequency) were related to (and hence predictable from) the background variables: (i) age at testing, (ii) current Indonesian/Dutch input, and (iii) current richness of Indonesian/Dutch input. If, for example, the scores on the dependent variables increase with age, we would expect a significant positive relation between age and the dependent variables. To this end, it is important to treat the predictors as continuous variables, rather than creating arbitrarily defined groups (e.g. age groups). In addition, we investigated whether macrostructure scores in the ML were related to length of ML exposure. In these analyses, Participant and Story (Cat; Dog; Baby Birds; Baby Goats) were

included in the random part of the model: in this way we took differences in general proficiency of the participants and possible differences in the difficulty of the stories (task effects) into account.

To address the second objective of this research, we created models that computed whether there was a relation between ML and HL scores for the three dependent variables: story structure, structural complexity, and number of ISTs. If there is cross-language transfer of narrative skills, we would expect a significant positive relation between the corresponding ML and HL scores. In these analyses, data was pooled over stories: the average score on the two stories in both languages was taken as the dependent measure. If there was a significant HL-ML relationship, we further investigated if this relationship was influenced by age and length of exposure to Dutch by adding these variables one by one to the model predicting ML scores from their HL counterparts. In this way, we could investigate whether the relation between the scores in Indonesian and Dutch was (in part) dependent on one of these background variables.

5. Results

5.1 Macrostructure scores in the two languages

The primary aim of this research was to predict narrative performance *within* languages (from age and exposure factors) and to study cross-language *relationships* (rather than differences). Before addressing these research questions we present the macrostructure scores to give an idea of the relative performance in the two languages. Table 2 presents descriptive statistics for story structure, story complexity and IST use in Indonesian (HL) and Dutch (ML). The macrostructure scores for story structure and story complexity were higher in Indonesian than in Dutch (t_{ss} (31) = −3.48, p = .002; t_{sc} (31) = −7.26, p < .001). There was no difference between the two languages in the frequency of ISTs (t(31) = −1.77, p = .087). In Indonesian, 22 children had mean story structure scores above 8, which is an expected score given the age of the participants (cf. Gagarina, 2016). However, in Dutch only 13 children had mean story structure scores above 8.

Table 3 presents percentages of different types of macrostructural sequences out of all opportunities to produce a sequence (3 opportunities per narrative * 2 narratives per language * 32 children = 192 opportunities per language). As can be seen in Table 3, more than half of the episodes in the Indonesian narratives were complete episodes, in stark contrast to 12% of GAO sequences in the Dutch narratives. Of 32 children in our sample, only 15 children had at least one complete episode (GAO) in Dutch, whereas all but one participant produced at least one

Table 2. Macrostructure production scores in Indonesian and Dutch (ranges represent mean values for two narratives produced per language)

	Indonesian (HL)		Dutch (ML)	
	Mean (SD)	**Range**	**Mean (SD)**	**Range**
Story structure	8.91 (2.52)	3–13.5	7.61 (2.33)	2–12
Story complexity	6.84 (1.80)	2–9	4.34 (1.70)	0.5–7
IST	4.44 (2.00)	1.5–9.5	3.75 (2.05)	0–8.5

GAO in Indonesian. The most prevalent types of sequences in the Dutch narratives were the less complex A-O sequences and no sequences (A or O).

Table 3. Percentages of different types of macrostructural sequences, by language

Types of sequences	Indonesian (HL)	Dutch (ML)
No sequence	6	6
Attempt/Outcome	7	28
Goal	5	8
Attempt-Outcome	7	35
Goal-Attempt/Goal-Outcome	15	11
Goal-Attempt-Outcome	60	12

In sum, the participants produced more complex narratives in the HL, which is compatible with the observation that Indonesian was not only (chronologically) the first, but also the primary language for most of the children in our sample. At the same time, the higher scores in Indonesian could also be due to task effects. Even though MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Goats) are parallel in terms of episodic structure, they have different plotlines and a different number of characters (3 in MAIN1 *vs.* 5 in MAIN2). Hence, the picture narratives used for Dutch might have been more complex (see, however, the discussion of earlier studies in Section 6.1).

5.2 Predictors of macrostructure scores in HL and ML

To address our first goal, we investigated whether the macrostructure scores were related to age. For the Indonesian stories this was the case for story structure ($\beta = 0.044$, $SE = 0.016$, $t = 2.68$, $p = .01$), but not for IST use ($\beta = 0.023$, $SE = 0.014$, $t = 1.63$, $p = .12$) and episodic complexity ($\beta = 0.020$, $SE = 0.013$, $t = 1.58$, $p = .13$). Thus, the older the children were, the more story grammar elements they pro-

duced in their HL narratives (see Figure 1). In Dutch, there was no relation with age for any of the measures (story structure: $\beta = 0.025$, $SE = 0.016$, $t = 1.58$ $p = .12$; structural complexity: $\beta = 0.013$, $SE = 0.012$, $t = 1.02$, $p = .32$; IST: $\beta = 0.013$, $SE = 0.015$, $t = 0.93$, $p = .36$).

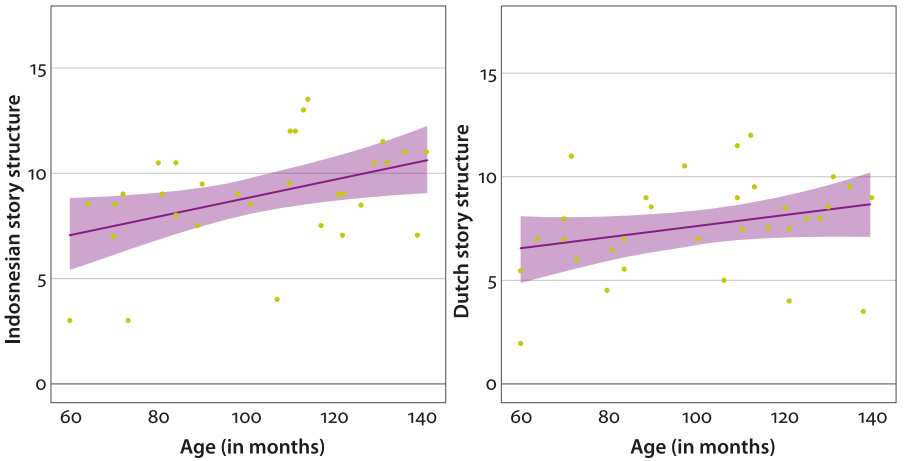

Figure 1. HL (Indonesian) and ML (Dutch) story structure scores predicted by age (datapoints are means of two narratives told in each language)

Next, we analysed the relationship between the HL and ML input measures and the macrostructure outcomes in the HL and ML respectively. Current Indonesian input did not predict any of the Indonesian macrostructure measures (story structure: $\beta = -0.048$, $SE = 0.081$, $t = -0.59$, $p = .56$; structural complexity: $\beta = 0.007$, $SE = 0.058$, $t = 0.13$, $p = .90$; IST: $\beta = 0.011$, $SE = 0.064$, $t = 0.17$, $p = .87$) and current Dutch input did not predict any of the Dutch macrostructure measures (story structure: $\beta = 0.007$, $SE = 0.086$, $t = 0.09$, $p = .93$; structural complexity: $\beta = 0.025$, $SE = 0.065$, $t = 0.38$, $p = .71$; IST: $\beta = 0.052$, $SE = 0.075$, $t = 0.70$, $p = .49$). The same holds for richness of Indonesian/Dutch input (HL story structure: $\beta = -0.239$, $SE = 0.297$, $t = -0.80$, $p = .43$; HL structural complexity: $\beta = 0.043$, $SE = 0.214$, $t = 0.20$, $p = .84$; IST use in HL: $\beta = 0.021$, $SE = 0.238$, $t = 0.09$, $p = .93$; ML story structure: $\beta = 0.283$, $SE = 0.348$, $t = 0.81$, $p = .42$; ML structural complexity: $\beta = 0.265$, $SE = 0.262$, $t = 1.01$, $p = .32$; IST use in ML: $\beta = 0.373$, $SE = 0.303$, $t = 1.23$, $p = .23$). Similarly, length of exposure to Dutch did not predict any of the Dutch macrostructure measures (story structure: $\beta = 0.008$, $SE = 0.010$, $t = 0.73$, $p = .47$; structural complexity: $\beta = 0.004$, $SE = 0.008$, $t = 0.54$, p = .59; IST: $\beta = 0.008$, $SE = 0.009$, $t = 0.86$, $p = .40$).

5.3 Relationship between macrostructure in HL and ML

ML story structure scores could be predicted from HL story structure ($\beta = 0.576$, $SE = 0.132$; $t = 4.38$, $p < .001$), see Figure 2. The positive estimate (β) indicates that the more story structure elements children produced in the HL, the more story structure elements they produced in the ML. Age did not affect this relationship ($\beta = 0.003$, $SE = 0.006$, $t = 0.54$, $p = .60$). However, the relationship between HL and ML became stronger as a function of length of exposure to ML ($\beta = 0.014$, $SE = 0.004$; $t = 3.27$, $p = .003$), even when controlled for Age ($\beta = 0.014$, $SE = 0.004$, $t = 3.21$, $p = .003$): The longer the exposure to Dutch, the stronger the relation between Indonesian and Dutch scores.

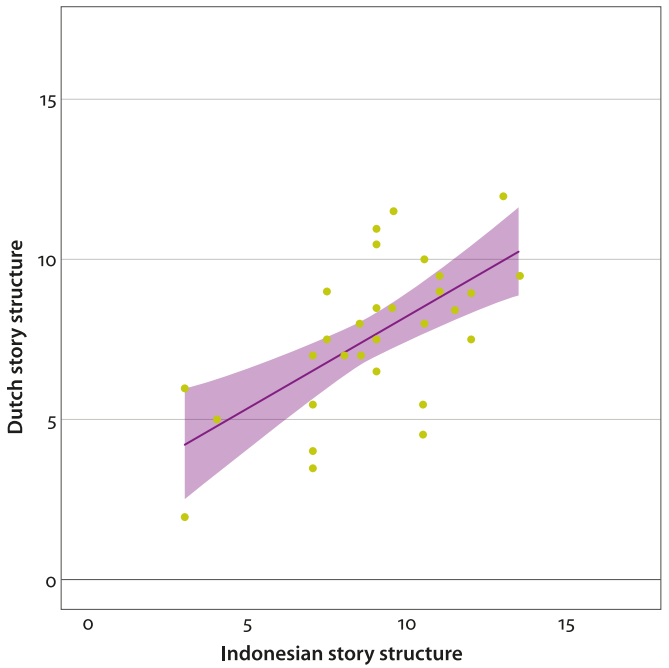

Figure 2. The relationship between story structure scores in HL (Indonesian) and ML (Dutch) (datapoints are means of two narratives told in each language)

There was also a significant positive relationship between structural complexity scores in HL and ML ($\beta = 0.361$, $SE = 0.160$; $t = 2.26$, $p = .03$), see Figure 3. The higher the score was in the HL, the higher it was in the ML. Age did not affect this relationship ($\beta = 0.007$, $SE = 0.006$, $t = 0.13$, $p = .27$). The relationship between HL and ML did become stronger as a function of length of exposure to ML ($\beta = 0.011$, $SE = 0.005$; $t = 2.13$, $p = .04$). However, this interaction was not conventionally sig-

nificant when controlled for Age ($\beta = 0.011$, $SE = 0.005$, $t = 2.05$, $p = .051$). Thus, the longer children had been exposed to Dutch, the stronger the relation between the structural complexity scores in Indonesian and Dutch was, but this effect was strongly correlated with Age.

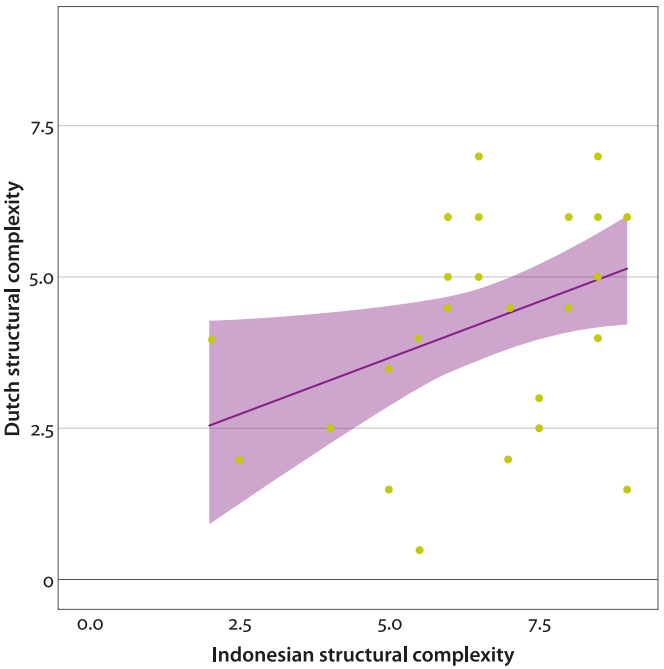

Figure 3. The relationship between structural complexity in HL (Indonesian) and ML (Dutch) narratives (datapoints are means of two narratives told in each language)

There was also a positive relation between IST frequencies in HL and ML ($\beta = 0.420$, $SE = 0.171$, $t = 2.46$, $p = .02$). The more ISTs children produced in their Indonesian narratives, the more ISTs they produced in their Dutch narratives (Figure 4). This relationship did not become stronger with length of exposure to Dutch ($\beta = 0.004$, $SE = 0.006$, $t = 0.63$, $p = .54$). Interestingly, the relationship became weaker with age ($\beta = -0.020$, $SE = 0.006$, $t = -3.22$, $p = 0.003$), even when controlled for length of exposure to Dutch ($\beta = -0.020$, $SE = 0.007$, $t = -2.93$, $p = .007$). Thus, the older the children were, the weaker the relation between HL and ML was.

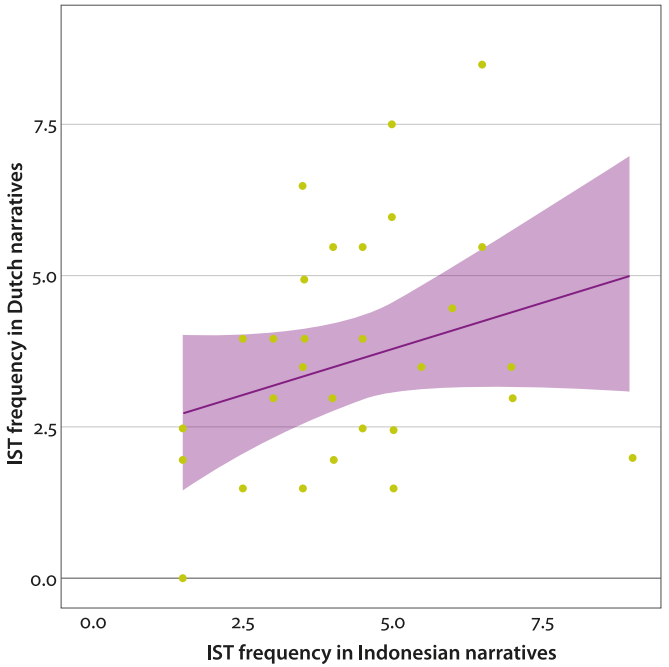

Figure 4. The relationship between IST use in HL (Indonesian) and ML (Dutch) (datapoints are means of two narratives told in each language)

6. Discussion and conclusion

Previous studies of narrative competence in bilingual populations have yielded controversial findings. Some report similar narrative skills in the two languages of bilingual children, which has been taken as evidence of positive transfer, whereas other studies find stronger narrative skills in one of the languages. Likewise, some studies report positive correlations between macrostructure scores in the two languages of a bilingual child (e.g., Lindgren, 2018; Öztekin, 2019), whereas other studies suggest that such cross-language relationships may be constrained by age and vary per narrative skill (e.g., Gagarina, 2016). This study set out to explore the relationship between narrative macrostructure in the ML and the HL, as well as the relation between narrative skills, on the one hand, and age and exposure, on the other hand. To this end, we analysed story structure, structural complexity and IST use in the narratives of children from Indonesian-speaking homes acquiring Dutch as a ML in the Netherlands.

6.1 The quality of HL and ML narratives

Even though the aim of this research was not to compare the performance in HL and ML directly, descriptive statistics presented in Section 5.1 have revealed higher story structure and structural complexity scores in Indonesian than in Dutch. The story structure scores and the proportion of full episodes in Indonesian were age-appropriate and similar to the L1 scores reported for older children in prior research (e.g. third-graders in Gagarina, 2016). In Dutch, the story structure scores were lower, and the types of sequences less complex (attempts and outcomes, few goals).

This difference is consistent with the finding that the participants (as a group) have had more cumulative exposure to Indonesian than to Dutch (two-thirds of the participants were born in Indonesia and spent their first years of life there) and with the parental judgement that the majority of the participants felt more comfortable using Indonesian. However, the difference in the scores can also be due to differences in complexity between the stories used in Indonesian (Cat/Dog) and Dutch (Baby Birds/Goats). Lindgren (2018) reports that in all three languages investigated in her study (Turkish, German and Swedish), the children scored higher on the *comprehension* of MAIN1 (Cat/Dog) compared to MAIN2 (Baby Birds/Goats), which suggests that the former set of narratives may indeed be less complex. At the same time, Lindgren found no task effects on the *production* of macrostructure. She suggests the possibility that "the comprehension questions of Cat/Dog are easier than those of Baby Birds/Goats, whereas the actual story content has a similar level of difficulty" (Lindgren, 2018, p. 219). In the same vein, Öztekin (2019) only found better performance on MAIN1 in comprehension, but not in production. Even though these previous studies have not found task effects that could explain better performance in Indonesian (MAIN1) than in Dutch (MAIN2), a more rigorous design allowing cross-language comparisons should involve full counterbalancing of the stories across languages.

Also, the order of testing might have influenced the performance in the two languages. For the reasons explained in Section 4.2.2, all children were tested first in Indonesian (Cat/Dog) and then in Dutch (Baby Birds/Goats). The same permanent order (Cat/Dog followed by Baby Birds/Goats) was also employed by Lindgren (2018). On the one hand, this order is not problematic in terms of possible carry-over effects, because such effects would lead to better performance in the language tested second (i.e. Dutch, which is not the case). Even if such learning effects occurred, they were not powerful enough to make the performance in Dutch similar to the performance in Indonesian. On the other hand, there might have been a decrease in performance due to fatigue, which could also partly explain the refusal of a number of (younger) participants to tell the stories in

Dutch. For practical reasons, both sessions were conducted on the same day by the same experimenter, which might also have influenced performance in Dutch. Future research targeting this population should administer the tests in the two languages on different occasions, by different experimenters and counterbalance the order.

Even though comparisons between ML and HL seem to reveal interesting patterns that need to be properly addressed in future studies, the primary aim of our research was to predict narrative performance *within* languages (from age and exposure factors) and to study cross-language *relationships* (rather than differences). Positive cross-language relationships entail that children who produce better stories in one language are also the ones who are likely to produce better stories in the other languages. In the remainder of this section we first discuss the results for the predictors of narrative quality and then turn to the relations between macrostructure in Indonesian and Dutch narratives.

6.2 The role of age and exposure

We predicted that age would be positively associated with all aspects of macrostructure in Indonesian (the language all participants acquired from birth), but not in Dutch, which was (chronologically) a second language for the participants in our sample. As expected, age at testing predicted story structure scores in Indonesian: The older the children were, the more story grammar elements their narratives contained. This finding supports previous research reporting age-related increases in story structure scores (Lindgren, 2018; Öztekin, 2019). This improvement is probably related to cognitive development and the development of theory of mind in particular (Wellman, Cross & Watson, 2001). Older children are better able to understand goal-driven behaviour of protagonists and therefore increasingly include goals in their narratives as a function of age (Trabasso & Nickels, 1992). Older children also realize that the listener does not necessarily have the same knowledge as they do (which is crucial because the experimenter could not see the pictures). Furthermore, older children have more experience with storytelling and more metacognitive knowledge of the elements that constitute good stories. Finally, older children have more linguistic resources at their disposal, which enables them to talk about what happened in the narrative and why it happened.

It is surprising that no relationship with age was found for structural complexity and IST use. Previous research has repeatedly demonstrated age-related improvements for both episodical complexity (e.g., Gagarina, 2016) and internal state talk (e.g., Chen & Yan, 2011; Ukrainetz et al., 2005). One methodological difference with prior research is that prior studies conducted group comparisons

(e.g., pre-schoolers *vs.* first graders), whereas this study predicted narrative performance from age used as a continuous variable. Since the age range in this research was relatively broad (5–11 years) and 12 participants had to be excluded because they refused to produce narratives in Dutch, the final sample might have been insufficient and too heterogenous to capture age-related improvements in structural complexity and IST frequencies. Future research will benefit from larger samples and more homogeneous samples (as in Govindarajan & Paradis, 2019).

As predicted, age was only related to narrative performance in Indonesian, but not in Dutch. The differential relationship between age and performance in the HL and the ML is probably related to greater variation in ML onset and proficiency. All children in this study were exposed to Indonesian from birth; their first exposure to Dutch varied between 5 months and 9 years. Hence, for Dutch older age does not necessarily mean better language skills. For example, Participant 039 is one of the youngest participants in this study (age 5;0), but she was born in the Netherlands and was exposed to Dutch from the age of five months. In contrast, Participant 040 is older (age 10;9), but was born in Indonesia and only arrived in the Netherlands two years prior to this study. In older L2 learners age and exposure interact in more complex ways than in L1 learners. On the one hand, a later age of onset places them at a disadvantage in terms of length/amount of L2 exposure. Even if the child is perfectly able to tell an elaborate narrative in the L1, limited L2 proficiency will not allow her to produce a narrative of the same quality in the L2. Accordingly, the Indonesian narrative of Participant 040 has a more elaborate story structure and episodic complexity than the Dutch narrative of the same participant (score 11 in Indonesian *vs.* 7 in Dutch, see Appendix). On the other hand, older children acquire an L2 faster due to their advanced cognitive and (meta)linguistic skills (Blom & Bosma, 2016; Blom & Paradis, 2015; Goldberg, Paradis, & Crago, 2008). Revisiting Participant 040, we see that her performance in L2 Dutch is better than what can be expected from a younger child after two years of exposure. The Dutch narrative of Participant 040 (provided in the Appendix) is much more elaborate than that of the younger participant (P039) with longer exposure to Dutch, see (1):

(1) (P039 [5;0], Baby Birds, Dutch, MAIN structure score = 2/17)
De kat. De kat, babyvogel, mamavogel. De mamavogel. En toen was daar de kat. De kat klimmen. In de boom. En dan daar is de hond. Toen de hond pakt de poes.
'The cat. The cat, baby-bird, mother-bird. The mother-bird. And then there was a cat. The cat climb. In the tree. And there is the dog. Then the dog grabs the cat.'

Since biological age is an index of both cognitive status and L1 proficiency, future research should try to tease the two factors apart. It is possible that language proficiency and cognitive status interact in non-trivial ways, especially in the development of narrative competence in the weaker language.

This study found no significant relationship between the macrostructure scores and the three exposure measures (length of exposure to Dutch, current Indonesian/Dutch input, current richness of Indonesian/Dutch input). This finding is in line with studies demonstrating that narrative quality is not predicted by length of L2 exposure (Schwartz & Shaul, 2013) and L1/L2 use in the home (Govindarajan & Paradis, 2019; Hipfner-Boucher et al., 2015). At the same time, the current results run counter to the finding that richness of input is positively related to narrative outcomes (Govindarajan & Paradis, 2019). However, Govindarajan and Paradis (2019) used a more homogenous sample of L2 learners, whereas our study had a relatively small sample with a wide range of variation in the exposure measures. It is also possible that our measure of richness was not sensitive enough to capture the influence of richness. The richness score used in this research was a sum of the number of speakers speaking the language with the child (at home) and the number of activities. We only counted the speakers that are included in the PaBIQ (mother, father, siblings, another adult, friends and family friends), but we do not know with how many other people the child communicated in that language (e.g., teachers and classmates), which is particularly problematic for studying richness of exposure to the ML. Similarly, the activities only included those covered by the questionnaire (reading, watching television/movies, storytelling), but not such activities as sports, music lessons and computer games. We credited one point for each speaker and each activity if the answer to that question was positive (e.g., if the activity belonged to the child's weekly life), irrespective of the frequency. As a result, many participants approached the maximum richness score for both languages (max = 9) and the variance in the sample was relatively low. Future studies would benefit from collecting more detailed information on the number of speakers providing input in each language and the variety of activities undertaken in that language.

6.3 Relationship between HL and ML narrative skills

The second objective of this study was to investigate the relationship between HL and ML across three different aspects of macrostructure (story structure, structural complexity and mental state talk). Positive cross-language relationships were found for all three macrostructural dimensions. Interestingly, these relationships were differentially predicted by age and exposure measures: For story structure and story complexity HL-ML relationships became stronger as a function of

length of exposure to the ML (but not age). And for IST use the opposite pattern emerged: The cross-language relationship became weaker with age (but did not depend on length of exposure to Dutch). Based on these findings, we concur with Gagarina (2016, p. 112) that "the various constituents of narrative skills such as story structure, story complexity, and ISTs do not exhibit an identical developmental trajectory, suggesting the 'fine-grained' composition of narrative skills". Prior studies have sometimes included different aspects of macrostructure into one "narrative quality" score (e.g., Squires et al., 2014; Uccelli & Paéz, 2007). Also, IST frequencies have commonly been included into more general expressive/subjective language scores, along with hedges, evaluative adjectives and causal connectives (Chen & Yan, 2011; Ukrainetz et al., 2005). The finding that story structure, structural complexity and mental state talk reveal different developmental trajectories, both in terms of cross-linguistic relationships and regarding the influence of age and exposure factors, warrants more research tracing the development of different narrative subskills separately and in comparison to one another.

The finding that the cross-language relations in story structure and episodic complexity became stronger as a function of length of exposure to the ML suggests that bilingual children with longer exposure to Dutch showed more transfer of story grammar elements from the HL. This result is consistent with our predictions. At early stages of L2 acquisition the child's L2 proficiency is probably too limited to allow positive transfer of narrative skills. If the child knows how to produce a good narrative in her stronger language, but does not know enough words and grammar to produce a story in the other language, usefulness of such narrative metaknowledge is likely to be very limited. With increasing exposure, children acquire more formal means that enable them to produce coherent stories in the L2; this is when their L1 knowledge of how to tell stories becomes relevant.

A different pattern emerged for mental state terms. In this case, it was the children's age, but not length of exposure to Dutch that influenced the relationship between HL and ML. We predicted that the cross-linguistic relationships would become stronger with age. This is because older children have a more developed theory of mind and are thus better able to understand protagonists' internal states. However, contrary to our predictions, the HL-ML relationship became *weaker* with age, even when controlled for length of exposure to Dutch. One explanation of this result might be that Dutch and Indonesian are typologically different languages. Even though Indonesian uses some Dutch loanwords, there is not much lexical overlap in the IST terms; therefore, the amount of positive lexical transfer is probably limited (cf. Blom et al., 2020).

At the same time, the influence of cognitive factors, theory of mind in particular, may become weaker with age. Important developmental changes in theory

of mind take place in pre-school and early school years (Wellman et al., 2001). After that, the differences between children level off. Theory-of-mind skills are related to understanding internal states in stories and to the production of ISTs (Curenton, 2004). Since differences in theory-of-mind capacities level off as a function of age, the impact of theory of mind on the use of ISTs is likely to abate. This assumption is supported by Greenhalgh & Strong (2001), a study that found no age-related increases in the use of ISTs in school-aged children (range 7–10 years old). In order to verify this idea, future work should relate the changes in cross-lingual relationships (in IST use) to changes in theory of mind.

Finally, as suggested by an anonymous reviewer, the negative relationship with age might be due to a gradual functional shift in language dominance. Based on parental reports, we can assume that Indonesian was still the dominant language for most of our participants. However, it is well-established that heritage children gradually become more dominant in their ML so that their L2 becomes their primary language (Montrul, 2010, 2012; Moro, 2016; Polinsky, 2008). Longitudinal studies tracing the development of HL and ML narrative competence, in relation to age and exposure measures, are warranted in order to get insights into such complex dynamic relationships.

It is important to notice that significant cross-language relationships say nothing about directionality: Transfer of narrative skills is likely to be from the stronger language to the weaker language. However, transfer from the weaker language is also possible, for example, if the child receives narrative training in the ML (Petersen et al., 2016). So it is possible that at least for some of the (older) children in our sample, transfer was from Dutch to Indonesian, especially given the fact that the participants were reported to have more experience with books and storytelling in Dutch than in Indonesian. Of 32 participants, 31 had engaged with Dutch books and 28 with storytelling in Dutch; for Indonesian this was only the case for half of the children (14 and 19 respectively). Longitudinal studies would be a promising avenue for future research that would allow to test predictions about the direction of transfer.

The present study only focused on the relationship between narrative (sub)skills with age and exposure, and did not look at the role of language proficiency. It is plausible to assume that proficiency in HL and ML could also influence cross-language relationships in narrative competence. Recall that studies with relatively balanced bilinguals (e.g., Fiestas & Peña, 2004; Kunnari et al., 2016) find similar narrative performance across the two languages of a bilingual. In contrast, studies with unbalanced bilinguals tend to find less complex narratives in the weaker language (e.g., Kapalková et al., 2016). According to Cummins (1979), only bilinguals attaining a certain threshold of proficiency in both languages experience cognitive and metalinguistic advantages of bilingualism. Future

research will benefit from studies that will empirically establish proficiency thresholds making narrative skills more amenable to cross-lingual transfer.

Funding

This research was supported by the Indonesia Endowment Fund for Education (LPDP RI), scholarship nr. S-210/LPDP.3/2017 to the second author.

Acknowledgements

We are very grateful to the three anonymous reviewers and the editors of this special issue for their detailed constructive comments on the earlier versions of the manuscript. We also would like to thank all participating families for making this research possible. Nienke Alblas kindly helped with the transcription and scoring of the Dutch narratives.

References

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–8. https://doi.org/10.18637/jss.v067.i01

Bishop, D.V.M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*(2), 156–173. https://doi.org/10.1044/jshd.5202.156

Blom, E., Boerma, T., Bosma, E., Cornips, L., Van den Heuij, K., & Timmermeister, M. (2020). Cross-language distance influences vocabulary outcomes of bilingual children. *First Language*, *40*(2), 151–171. https://doi.org/10.1177/0142723719892794

Blom, E., & Bosma, E. (2016). The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian-Dutch children. *Journal of Child Language*, *43*(3), 581–607. https://doi.org/10.1017/S0305000915000574

Blom, E., & Paradis, J. (2015). Sources of individual differences in the acquisition of tense inflection by English second language learners with and without specific language impairment. *Applied Psycholinguistics*, *36*(4), 953–976. https://doi.org/10.1017/S014271641300057X

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language and Communication Disorders*, *51*(6), 626–638. https://doi.org/10.1111/1460-6984.12234

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

Chen, L., & Yan, R. (2011). Development and use of English evaluative expressions in narratives of Chinese-English bilinguals. *Bilingualism: Language and Cognition*, *14*(4), 570–578. https://doi.org/10.1017/S1366728910000362

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning with specific language impairment. *Journal of Communication Disorders*, *43*(6), 511–522. https://doi.org/10.1016/j.jcomdis.2010.05.005

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Curenton, S. M. (2004). The association between narratives and theory of mind for low-income pre-schoolers. *Early Education and Development*, *15*(2), 124–146. https://doi.org/10.1207/s15566935eed1502_1

Curenton, S. M. (2011). Understanding the landscapes of stories: The association between preschoolers' narrative comprehension and production skills and cognitive abilities. *Early Child Development and Care*, *181*(6), 791–808. https://doi.org/10.1080/03004430.2010.490946

Curenton, S. M., & Justice, L. M. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools*, *35*(3), 240–253. https://doi.org/10.1044/0161-1461(2004/023)

De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol, etc.: Multilingual Matters. https://doi.org/10.21832/9781847691507

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

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

Fusté-Herrmann, B., Silliman, E. R., Bahr, R. H., Fasnacht, K. S., & Federico, J. E. (2006). Mental state verb production in the oral narratives of English- and Spanish-speaking preadolescents: An exploratory study of lexical diversity and depth. *Learning Disabilities Research and Practice*, *21*(1), 44–60. https://doi.org/10.1111/j.1540-5826.2006.00206.x

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

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Valimaa, T., Balciuniene, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual assessment instrument for narratives. *ZAS papers in linguistics 56*: http://www.zas.gwz-berlin.de/zaspil56.html

Goldberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority L1 children learning English as a L2. *Applied Psycholinguistics*, *29*(1), 1–25.

Govindarajan, K., & Paradis, J. (2019). Narrative abilities of bilingual children with and without Developmental Language Disorder (SLI): Differentiation and the role of age and input factors. *Journal of Communication Disorders*, *77*, 1–16. https://doi.org/10.1016/j.jcomdis.2018.10.001

Greenhalgh, K. S., & Strong, C. J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools*, *32*(2), 114–125. https://doi.org/10.1044/0161-1461(2001/010)

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

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, *19*(6), 677–692. https://doi.org/10.1177/1367006914534330

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

Klamer, M., & Moro, F. (2020). What is "natural" speech? Comparing free narratives and Frog stories in Indonesia. *Language Documentation and Conservation*, *14*, 238–313.

Krasnoshchekova, S., & Kashleva, K. (2019). Narrative competence of adult L2 Russian learners. *Journal of Psycholinguistic Research*, *48*(1), 617–641. https://doi.org/10.1007/s10936-018-9622-3

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

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6*. (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McLaughlin, B. (1978). *Second language acquisition in childhood*. Hillsdale, NJ: Lawrence Erlbaum.

Montanari, S. (2004). The development of narrative competence in the L1 and L2 of Spanish-English bilingual children. *International Journal of Bilingualism*, *8*(4), 449–497. https://doi.org/10.1177/13670069040080040301

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, *30*, 3–23. https://doi.org/10.1017/S0267190510000103

Montrul, S. (2012). Is the heritage language like a second language?. *Eurosla Yearbook*, *12*(1), 1–29. https://doi.org/10.1075/eurosla.12.03mon

Moro, F. R. (2016). Dynamics of Ambon Malay: Comparing Ambon and the Netherlands. (LOT Dissertation Series 422). Utrecht: LOT.

Öztekin, B. (2019). *Typical and atypical language development in Turkish-Swedish bilingual children aged 4–7* (Studia Linguistica Upsaliensia 25). Uppsala: Acta Universitatis Upsaliensis.

Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: Individual variation in long-term outcomes. *Journal of Child Language*, *43*(3), 553–580. https://doi.org/10.1017/S0305000915000562

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721-008

Petersen, D. B., Thomsen, B., Guiberson, M. M., & Spencer, T. D. (2016). Cross-linguistic interactions from second language to first language as the result of individualized narrative language intervention with children with and without language impairment. *Applied Psycholinguistics*, *37*(3), 703–724. https://doi.org/10.1017/S0142716415000211

Polinsky, M. (2008). Heritage language narratives. In D. Brinton, O. Kagan & S. Bauckus (Eds.), *Heritage language education: A new field emerging* (pp. 149–164). New York: Routledge.

Schwartz, M., & Shaul, Y. (2013). Narrative development among language-minority children: The role of bilingual versus monolingual preschool education. *Language, Culture and Curriculum*, *26*(1), 36–51. https://doi.org/10.1080/07908318.2012.760568

Silliman, E. R., Bahr, R. H., Brea, M. R., Hnath-Chisolm, T., & Mahecha, N. R. (2002). Spanish and English proficiency in the linguistic encoding of mental states in narrative retellings. *Linguistics and Education*, *13*(2), 199–234. https://doi.org/10.1016/S0898-5898(01)00062-6

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

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, *23*(1), 236–255. https://doi.org/10.1177/1367006917722418

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*(3), 249–275. https://doi.org/10.1080/01638539209544812

Tribushinina, E., Mak, W. M., Andreiushina, E., Dubinkina, E., & Sanders, T. (2017). Connective use by bilinguals and monolinguals with SLI. *Bilingualism: Language and Cognition*, *20*(1), 98–113. https://doi.org/10.1017/S1366728915000577

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–121). Bristol, etc.: Multilingual Matters. https://doi.org/10.21832/9781783093137-013

Uccelli, P., & Páez, M. 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*(3), 225–236. https://doi.org/10.1044/0161-1461(2007/024)

Ukrainetz, T. A., Justice, L. M., Kaderavek, J. N., Eisenberg, S. L., Gillam, R. B., & Harm, H. M. (2005). The development of expressive elaboration in fictional narratives. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1363–1377. https://doi.org/10.1044/1092-4388(2005/095)

Unsworth, S. (2005). Child L2, adult L2, child L1: Differences and similarities. A Study on the acquisition of direct object scrambling in Dutch. PhD thesis, Utrecht University.

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*, 86–110. https://doi.org/10.1017/S1366728912000284

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Appendix. Sample transcriptions, coded for story structure

IST-IE = internal state term as initiating event; G = goal; A = Attempt; O = outcome; IST-R = internal state term as reaction

P040 [age 10;9], Dog, Indonesian (MAIN structure score = 11/17)

*CHI: eh ada Anjing sama Tikus.
'Oh there is a dog and a mouse.'
*CHI: ada Pohon juga.
'There is also a tree.'
*CHI: terus Anjingnya itu kepengen [: ingin][*] makan Tikusnya. [G]
'Then the dog wants to eat the mouse.' (A3)[4]
*CHI: terus Tikusnya lagi kedalam [x 2] Pohon. [O]
'Then the mouse is running into the tree.' (A5)
*CHI: di bawahnya Pohon gitu ada lubang.
'Under the tree there is a hole.'
*CHI: hla setelah masuk Anjingnya la [//] semacam lari gitu.[A]
'After he has entered it, the dog is sort of running.' (A4)
*CHI: terus ada Orang dikejar sampe [: sampai][*] situ.
'Then there came someone.'
*CHI: terus ada Orang pake [: membawa][*] Balon warna kuning.
'Then there is a person who is carrying a yellow balloon.'
*CHI: ndak@s:javanese tahu bawa Daging di Tangannya.
'I do not know if he has meat in his hands.'
*CHI: terus Dia pas@s:javanese mau dikejar ketubruk@s:javanese Pohon. [G, O]
'Then he wants to pursue him, but bumps into the tree.' (A3, A5)
*CHI: itu apa namanya Anjingnya.
'What is the name of the dog?'
*CHI: terus Tikusnya semacam ketawa gitu. [IST-R]
'Then the mouse is sort of laughing.' (A6)
*CHI: seneng [: senang][*] hla karna [: karena][*]. [IST-R]
'Happy because.' (A6)
*CHI: hla terus pas@s:javanese di situ Balonnya Dia yang Anaknya itu nggak [: tidak][*] engaja dilepaskan.
'Then the boy by accident released his balloon.'
*CHI: terus <nyanthol@s:javanese di Pohon> [x 2].
'Then the balloon is stuck in the tree.'
*CHI: hla pas@s:javanese Dia lagi lihat [: melihat][*] Pohonnya.
'Then he looks at the tree.'

4. These codes in round brackets represent the relevant parts of the MAIN scoring scheme.

*CHI: lagi lihat [: melihat][*] Balonnya. [IST-IE]
'He sees the balloon.' (A7)
*CHI: terus Anjingnya semacem [: semacam][*] lihat [: melihat][*] ke Dagingnya. [IST-IE]
'Then the dog sort of sees the meat.' (A12)
*CHI: ho ini Dagingnya enak gitu. [IST-IE]
'Hey, the meat looks delicious.' (A12)
*CHI: terus akhirnya Dia.
'Then finally he.'
*CHI: yang Anak kecilnya itu ndak@s:javanese [: tidak][*] tahu kalo [: kalau][*] ada Anjing yang mau makan itu. [IST-IE, G]
'The boy does not realize that there is a dog who wants to eat it.' (A13)
*CHI: jadi Dia tetep [: tetap][*] ngambil [: mengambil][*] Balonnya. [A]
'Thus he is still trying to grab his balloon.' (A9)
*CHI: akhirnya Dia loncat [: meloncat][*] gitu diambil Balonnya. [A, O]
'Finally he jumps and grabs the balloon.' (A10)
*CHI: hla pas@s:javanese diambil Balonnya Anjingnya sudah ngambil [: mengambil][*] satu. [A]
'When he grabs the balloon, the dog takes one (piece of meat).' (A10, A14)
*CHI: setelah An [///] setelah Dia sudah ambil [: mengambil][*] Balonnya dipegang Tangannya. [O]
'After he takes the balloon in his hand.' (A10)
*CHI: sudah dimakan. [O]
'He has already eaten it.' (A15)
*CHI: ndak@s:javanese [: tidak][*] semuanya satu dimakan. [O]
'Not all of them, only one is eaten.'

P040 [age 10;9], Baby Birds, Dutch (MAIN structure score = 7/17)

*CHI: er is er drie vogels.
'There is there three birds.'
*CHI: twee kleine vogeltjes en één grote.
'Two little birds and one big bird.'
*CHI: en dan die grote vliegt weg. [A]
'And then the big one flies away.' (A4)
*CHI: <en een kat kom> [///] er is een kat.
'And a cat come [///] there is a cat.'
*CHI: en die kat aanvalt [: valt aan][*]. [A]
'And that cat attacks.' (A9)
*CHI: die ging op het [: de][*] boom klimmen. [A]
'It started climbing the tree.' (A9)
*CHI: en de kat wil het [: de][*] vogels eten. [G]
'And the cat wants to eat the birds.' (A8)
*CHI: dus pakt die [x 2.] kleine vogeltje. [O]
'So he grabs that little birdie.' (A10)
*CHI: en dan komt een hond.
'And then a dog comes.'

*CHI: en die hond trekt [x 2.] die kat weg. [A]
'And that dog pulls the cat away.' (A14)
*CHI: en dan het [: de][*] vogels zijn veilig. [IST-R]
'And then the birds are safe.' (A16)
*CHI: en het en dan het die kat rent weg. [O]
'And it and then that cat runs away.' (A14)

Address for correspondence

Elena Tribushinina
Universiteit Utrecht
Trans 10
3512 JK Utrecht
Netherlands
e.tribushinina@uu.nl
https://orcid.org/0000-0003-1688-7307

Co-author information

Mila Irmawati
University of Indonesia
mila.irmawati@alumni.ui.ac.id

Pim Mak
Utrecht University
w.m.mak@uu.nl

Publication history

Date received: 25 March 2020
Date accepted: 17 December 2020
Published online: 16 February 2021