Macrostructure in narratives produced by Lebanese Arabic-French bilingual children

Developmental trends and links with language dominance, exposure to narratives and lexical skills

Rachel Fiani[1], Guillemette Henry[1], and Philippe Prévost[2]
[1] Saint Joseph University of Beirut | [2] UMR1253 iBrain University of Tours

Studies of macrostructure in bilingual children investigated potential age and language effects, without systematically taking into account bilingualism factors such as language dominance, exposure to narratives, and general language abilities. These issues were addressed in a study of macrostructure production by 69 bilingual Lebanese Arabic-French children aged 4–9. The children were administered the Multilingual Assessment Instrument for Narratives in the tell mode and a standardized conceptual vocabulary test. Parental questionnaires were also used to gather information on language exposure and use. Age effects and limited effects of language were found for all measures of macrostructure production (story structure, structural complexity and use of internal state terms). Significant correlations arose between these measures, vocabulary scores, and age, and between structural complexity scores and language dominance. Regression analyses revealed the predictive role of vocabulary in the development of macrostructure. These results are compatible with the idea that macrostructure is not language dependent. They also suggest that minimal language skills are needed for expressing macrostructure. The limited impact of bilingualism factors could stem from the fact that French and Lebanese Arabic are majority languages in Lebanon, meaning that the children had received continuous exposure to both languages, including within the school system.

Keywords: narrative macrostructure, bilingualism factors, vocabulary, Lebanese Arabic, French

https://doi.org/10.1075/lab.20035.fia | Published online: 7 September 2021
Linguistic Approaches to Bilingualism 12:4 (2022), pp. 446–478. ISSN 1879-9264 | E-ISSN 1879-9272

1. Introduction

Macrostructure is the narrative structure of a story. According to story grammar models, macrostructure typically involves elements pertaining to the background of the story (the setting, including time, place, and introduction of the characters) and one (or more) episode(s) each containing an initiating event which in turn sets up the characters' goals for solving the situation (their internal plans), their attempts at reaching these goals, the consequences or outcomes of these attempts, and the characters' reactions toward the outcomes (Stein & Glenn, 1979, among others). Macrostructure has been argued to be universal, meaning that the way a story is constructed shouldn't vary across languages and cultures, while the linguistic means used to tell the story (microstructure) may vary from individual to individual and across languages. Research on macrostructure development in bilingual children has been booming in the last few years, notably due to the development of the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters, 2012) during COST Action IS0804. MAIN was created in order to allow for maximum comparison across narrative datasets. However, despite the increasing number of studies on narrative abilities of bilingual children using MAIN, the impact of language exposure and exposure to narratives, as well as development beyond the age of seven needs further investigation, including in contexts of simultaneous bilingualism. The aim of this study is to explore these issues in narratives produced by Lebanese Arabic-French bilingual children living in Lebanon, a multilingual context seldom explored in research on bilingual development.

2. Development of macrostructure production bilingual children

2.1 Measuring macrostructural skills

Macrostructural skills can be measured by tracking the number of macrostructure elements, such as the ones mentioned above, that are present in a narrative produced by a child. The complexity of the story structure can also be taken into account for evaluating macrostructural abilities. In many studies, story structure complexity is evaluated based on Westby's (2012) classification of narrative sequences. Central to this model is whether or not the goal is part of narrative sequences, since the goal is considered to encapsulate the intentions of the protagonist and thus to be fundamental for the coherence of the story (Gagarina et al., 2012). According to Westby's model, narrative episodes containing no sequence (and no goal) are considered least complex (e.g., attempt only or outcome only).

Sequences of two elements are deemed more complex, but within these sequences, absence of a goal (e.g. attempt-outcome) is viewed as less complex than presence of a goal (e.g. goal-attempt and goal-outcome). Finally, a complete goal-attempt-outcome (GAO) sequence is considered to be the most complex structure.

Another way of capturing the development of macrostructure skills is to look at the production of internal state terms (ISTs), namely terms reflecting the characters' mindsets. These are instrumental in characterizing the protagonists' emotional or mental state at the beginning of a narrative episode, their goals, their attempts at reaching these goals, and their reactions toward the outcomes of their actions. ISTs include terms reflecting emotions, beliefs, motivation, and experience.

2.2 Macrostructure skills in bilingual children

In bilingual children, a significant age effect on the development of macrostructure skills has been reported in the literature. Importantly, most studies tend to focus on the 4–7 age range, including MAIN-based studies, with only a few studies including older children, although macrostructure has been shown to continue developing beyond the age of 10, in parallel with the complexification of their discourse (Berman & Slobin, 1994). Many studies on narratives produced by bilingual children have also reported remarkably similar performance in the children's languages on measures of macrostructure production, even in the case of sequential bilinguals (Fiestas & Peña, 2004). In particular, it has been reported that story structure scores in the two languages may be correlated at a given age, without there being any significant correlations between scores measuring proficiency in the two languages in question, suggesting that story structure develops globally, as a function of age, independently of language skills (Pearson, 2002).

Zooming in on those studies that investigated macrostructure in narratives generated by bilingual children using MAIN, an age effect and a limited effect of language was also reported, with both simultaneous and sequential bilinguals (for a systematic review, see Appendix A). However, despite similar tendencies, some differences can be found across studies as well. Moreover, studies differ as to which measures of structural complexity were used, and few of them looked at ISTs, hampering inter-study comparisons.

2.2.1 *Previous findings on story structure in MAIN studies*

Although an age effect is reported with respect to story structure in both of the child's languages in some studies, it is not in others (or the age effect appears to be stronger in one language than in the other). Investigating Swedish-English bilinguals (age 4–7), Bohnacker (2016) observed a significant age effect for story struc-

ture in each language (see also Gagarina, 2016, on Russian-German children). In contrast, Roch, Florit and Levorato (2016) found a limited age effect for story structure in L1 Italian-L2 English children (age 5–7). In particular, story structure scores were found to increase across age groups in the L2, but not in the L1. Lindgren (2018) reported an age effect for story structure in her groups of simultaneous German-Swedish bilinguals, and in her groups of simultaneous and sequential Turkish-Swedish bilingual children (all aged 4 to 6 years), in each of the children's languages. However, the scores in Turkish were found to stagnate (between ages 5 and 6), in contrast to the scores in German. These findings were confirmed in a larger study of Turkish-Swedish children, with scores on story structure increasing more rapidly with age in Swedish than in Turkish (Öztekin, 2019).

Performance on story structure was also compared across the children's languages. Some studies reported no significant language effect, for example Kunnari, Välimaa and Laukkanen-Nevala (2016), which investigated Finnish-Swedish simultaneous bilinguals aged 5 to 6. In contrast, Lindgren (2018) reported a language difference in her German-Swedish children, especially in the older group (age 6), in favor of Swedish; such a difference was not observed in her Turkish-Swedish children (see also Öztekin, 2019). In sequential bilinguals, Roch et al. (2016) found similar performance in the minority language (English) and the majority language (Italian) in the older children (age 6–7), but not in the younger ones (age 5–6) who produced more macrostructural elements in their L1 (Italian).

Finally, out of all macrostructure elements, attempts and outcomes have been found to be most commonly produced (Bohnacker 2016; Lindgren, 2018; Öztekin, 2019), which has also been observed with monolingual children (Lindgren, 2019). This could be due to the fact that attempts and outcomes are visually salient in the MAIN pictures, in contrast to goals (Lindgren, 2018, 2019). No age effect was reported for production of goals in Lindgren (2018) but an increase in the number of goals was noted in Kunnari et al. (2016).

2.2.2 *Previous findings on structural complexity in MAIN studies*

Studies based on MAIN also reported on the complexity of story structure, via different means. Some studies reported on the percentage of each narrative sequence produced out of the total number of all possible sequences. Kunnari et al. (2016) found that around 80% of the episodes contained at most one isolated (attempt or outcome) element in their data, with no language effect. GAO episodes, which are the most complex narrative sequences, represented at most 15% of all sequences produced. Similarly, in Öztekin (2019), isolated elements (i.e. a lack of narrative sequences) were very frequent in both Turkish and Swedish. At age 7, a lack of

sequences was still very common, with complete GAO episodes occurring very rarely (15% at most). In Bohnacker (2016) an age effect, but no language effect, was reported concerning the percentage of GAO sequences produced, which rose from 7% at age 5 to 12% at age 6–7 (in each language).

Other studies looked at the highest level of complexity found in a child's story, focusing on the percentage of children producing sequences of different levels of complexity, irrespective of the number of sequences produced for each type. Roch et al. (2016) reported that while story structure complexity was higher in the majority language (Italian) at an early age (5–6) than in the minority language (English), with children using fewer complete GAO episodes and more AO episodes in English than in Italian, it was comparable in the two languages in the older children (age 6–7). In particular, GAO sequences were produced by about 34% of the children aged 6–7 in either language. In contrast, in Gagarina's (2016) study, although there was an age effect concerning use of GAO for both languages, the effect was more marked in Russian (the minority language) than in German (the majority language). In the older group of children (age 7;11–10;6), more than 60% of the children produced a GAO sequence in Russian versus less than 40% for German.

The two ways of measuring structural complexity were reported by Lindgren (2018). Regarding the percentage of GAO sequences produced out of all possible sequences, an age effect was found (2%–5% at age 4 versus 8%–12% at age 6), but no language effect (see also Öztekin, 2019). However, when looking at the proportion of children producing at least one GAO sequence, there was a language effect in the Turkish-Swedish children (with twice as many children producing GAO sequences in Turkish than in Swedish – 31.3% versus 16.7%), but not in the German-Swedish children (21.7% for each language). Unfortunately, the latter findings were conflated across age groups, thus preventing further comparisons.

Finally, although MAIN allows for the calculation of a cumulative complexity score which takes into account the complexity score of each of the episodes told by the child (see below), it has rarely been used so far, which is quite surprising. Indeed, by providing a score ranging from 0 to 9 points, it allows for more variability and potentially more precision concerning macrostructural abilities than a score based on only one episode, i.e., the most complex episode produced by the child (0–3 points). One of the goals of our study was to compare the different scoring methods, including the cumulative complexity score.

2.2.3 *Previous findings on ISTs in MAIN studies*

Few MAIN-based studies looked at ISTs in their investigation of macrostructure skills. Kunnari et al. (2016) reported low use of ISTs, which was taken to reflect emerging capacities for expressing the mindset of others. An age effect was found

for each language in Roch et al.'s (2016) study (see also Chen & Yan, 2011, on Chinese-English bilinguals aged 5–10). As to potential language effects, none was reported in Kunnari et al. (2016), which looked at simultaneous bilinguals. In sequential bilinguals, a language effect was found in the younger children in Roch et al.'s (2016) study, in favor of the L1 (Italian), but no language effect was observed in the older children. In contrast, in Gagarina's (2016) study, no language effect was found in the younger children, but in the older groups of children significantly more ISTs were produced in German (the majority language) than in Russian (the minority language). Note that studies differ as to what was counted as ISTs: Roch et al. only considered ISTs when used as initiating events and reactions to the outcome, whereas Kunnari et al. and Gagarina took into account a wider set of ISTs, making conclusive comparisons difficult.

2.2.4 *Potential impact of bilingualism factors and language proficiency*

Likely explanatory factors for differences across studies include the types of bilingual contexts in which the participants were growing up, including differential exposure to, and use of the two languages, especially through schooling which in some cases concerns the majority language only and in others involves both languages. Naturally, bilingual contexts were taken into account by researchers, sometimes by comparing simultaneous and sequential bilinguals (see Gagarina, 2016), but studies usually did not operationalize variables related to bilingualism, such as language exposure, language use and age of onset, in their results. Variability being one of the hallmarks of bilingualism, thoroughly investigating the impact of bilingualism factors on narrative performance is essential. It also facilitates comparisons across studies. One of these factors is language dominance which can be seen as a construct reflecting language use and exposure. Language dominance is operationalized in different ways in the literature, integrating various parameters such as age of onset, length of exposure, contexts of exposure, and language use at home and outside of the home (Treffers-Daller, 2019). Likewise, in our study a language dominance index was calculated based on answers from a parental questionnaire documenting the child's language history, language exposure, language use and language skills (see below).

Potential effects of language exposure include exposure to narratives. Some studies on story production reported that the frequency of storytelling affects narrative skills in both monolingual children (Leseman, Scheele, Mayo & Messer, 2007) and bilinguals (Bitetti & Hammer, 2016). However, such an effect has not been thoroughly investigated in studies on macrostructural skills of bilinguals, including MAIN studies (but see Fiani, Henry & Prévost, 2020, who found significant correlations between comprehension of macrostructure and frequency of exposure to stories in LB-French bilinguals).

Another potential factor which is yet to be explored is language proficiency. Although language abilities are clearly involved in the expression of macrostructure, as alluded to by several studies (Bohnacker, 2016), few studies conducted systematic investigations of the relationship between language skills and development of macrostructure. Lindgren (2018) found an effect of lexical skills on macrostructure performance (see also Uccelli & Páez, 2007). This was the case for Swedish in the two bilingual groups that were investigated, as well as for Turkish, but not for German. However, Öztekin (2019) found no vocabulary effect in macrostructure production by Turkish-Swedish children, although some children had low performance in both areas.

3. The study

Our study focuses on macrostructural skills in LB-French bilinguals growing up in Lebanon. by looking at production, via MAIN. Several types of bilingualism exist in Lebanon where three main languages are most commonly used (Arabic, French and English), as well as several minority languages such as Armenian. However, multilingual profiles vary depending on the region, education, social and cultural backgrounds. LB is acquired from birth, and French is often introduced very early, either at home through family members, or at school (as of age 3). Proficiency in French may vary depending on the type of school (private/public for example) or the region (Hafez, 2006). French can also be frequently heard outside of the home, as about 50% of the population is estimated to speak French (Leclerc, 2015), at daycare or through the media, including television. Given the status of LB and French as majority languages in Lebanon, this situation differs from those found in most studies on MAIN published so far, which have largely focused on bilingual contexts where there is a dichotomy between the home language and the societal language.

Three measures of macrostructure (story structure, story structure complexity, and use of ISTs) were used in our study, which also investigated the impact of language dominance, exposure to narratives, conceptual vocabulary on these measures. The study was also meant to make a methodological contribution to the MAIN literature by comparing different ways of capturing macrostructural complexity. Our research questions were the following:

1. Is there an age effect and/or a language effect on the production of macrostructure in each language? Do different ways of capturing macrostructural complexity impact the results?

2. What is the impact of language dominance, exposure to narratives and language proficiency on macrostructure performance in each language?

4. Method

4.1 Participants

The study included 69 LB-French bilingual children aged 4 to 9. There were 34 boys and 35 girls who were recruited at six private schools where French is the language of instruction.[1] The children were divided into three age groups (4- to 5-year-olds, 6- to 7-year-olds, and 8- to 9-year-olds; henceforth G4–5, G6–7 and G8–9) (see Table 1). All children came from a middle socio-economic background, based on the parents' level of education and current occupation. They were mostly simultaneous bilinguals: all of them were exposed to LB and French from birth, except for two (in G6–7) who started being exposed to French at 30 and 36 months respectively. Finally, all children had typical language development, as attested by their performance on a standardized language battery (Evaluation of oral language – LB, ELO-L, Zebib, Henry, Khomsi, Messarra & Hreich, 2017), which was within norms.

4.2 Materials

4.2.1 *Conceptual vocabulary task (ELO-L)*

Conceptual vocabulary was tested via a lexical production task from the ELO-L battery. For this task, children were presented with pictures, representing either nouns or verbs, which they were asked to label. Answers provided in any language were acceptable.

4.2.2 *Parental questionnaire (PaBiQ)*

A questionnaire for parents of bilingual children (PaBiQ, Tuller, 2015) was used to gather information about the children's language history, exposure, and use, and about the parents' level of education. The parents of three children refused to answer this questionnaire, as well as a questionnaire on exposure to narratives (see below). Information obtained on age of onset, language exposure contexts before the age of four, current language skills, languages used at home and languages used outside the home during routine activities were used to calculate a

1. The schools were located in different parts of the country: two were in Beirut, two were in the North (Tripoli and Koura), and two were in the South (Saida and Tyre).

language richness index (LRI) for each language. A language dominance index (LDI) was then obtained by subtracting the LRI for French from the LRI for LB. A positive LDI indicated dominance in LB, whereas a negative LDI meant that the child was dominant in French.

4.2.3 *Exposure to narrative questionnaire*

Another parental questionnaire was used to gather information about the children's exposure to stories (Fiani et al., 2020). This questionnaire allowed us to obtain an exposure to narratives index (ENI) for each language. This index reflected the wealth of exposure to oral narratives before the age of four and currently (i.e. stories that were read to the children by their parents), as well as the wealth of exposure to written narratives (i.e. stories that children read by themselves), in LB and in French.

4.2.4 *MAIN*

In our study, LB and French versions of MAIN were used to elicit narratives, following a story generation format (Gagarina et al., 2012).[2] The Baby Birds story was used for LB and the Baby Goats story was used for French, based on results from a pilot study on 18 bilingual children.[3]

4.3 Procedure

Following approval by the Ethical Committee of Saint Joseph University of Beirut, schools were contacted, and a letter of consent was signed by the parents of the participants. Testing was conducted in a quiet room at the children's schools, while the parental questionnaires were administered over the phone. All children were seen for two separate sessions of 30 minutes each, with an interval of approximately one week. Two research assistants, both native speakers of LB and French, conducted the testing sessions. Each one assessed the child in a different language in order to establish a monolingual context. Thirty-five children started with the

2. The LB version used here was developed by Rachel Fiani and Guillemette Henry (Saint Joseph University of Beirut, Lebanon) and the French version was developed by Martin Haiden (University of Nantes, France) and Alfred Knapp (University of La Rochelle, France). For the adaptation process in LB, see Fiani et al. (2020).

3. In this pilot study a high level of code-switching was observed in the Baby Goats story when told in LB, which affected the analysis of story production in both languages. Therefore, in the present study, stories were not counterbalanced across the two languages (see the discussion on this limitation).

session in LB and 34 started with the session in French. No order effect was found in any of the three age groups.

During assessment, all guidelines for MAIN were respected (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters, 2015). The effect of shared knowledge was controlled for by not allowing the experimenter to look at the pictures with the child during story elicitation. Three different envelopes were presented to the child with the MAIN stories inside. All of them contained the same story. The child was asked to choose one envelope and to carefully look at the pictures found inside, without showing them to the experimenter. After the child had looked at the pictures, the experimenter folded them back in pairs while facing the child and without looking at them, and unfolded them progressively during assessment. Presenting the pictures two at a time was meant to facilitate the production of the three episodes in each story. During storytelling, the experimenter was not allowed to use any prompts, except if the child fell silent. In that case, the French and LB equivalents of 'Ok', 'Continue', etc. were used to motivate the child to go on.[4]

All narratives were audio and video recorded and were transcribed into CHAT format, following CHILDES guidelines (MacWhinney, 2000). They were then coded for macrostructure. Transcriptions and coding were verified by two independent researchers. All disagreements (related to IST classification or to macrostructure scoring for example) were solved via discussions.

4.4 MAIN Measures

MAIN includes a protocol for assessing production of macrostructural components and structural complexity, which was used in this study.

Story structure

The story structure measure accounts for the number of story structure elements used by a child in each of the three episodes: 1 point was awarded for each of the five following story components for each episode: goal of the main protagonist, attempt, outcome, and internal states as initiating event and reaction. These made up a maximum of 15 points for all three episodes. In addition, 2 points were awarded for including the setting (1 point for time reference and 1 point for place reference), for a grand total of 17 points for the whole story.

4. After the child had finished telling the story, the experimenter proceeded to ask comprehension questions, the results of which were presented elsewhere (Fiani et al., 2020).

Story structure complexity

In our study, three different means of capturing structural complexity were compared. First, a structural complexity score was computed for every episode of a story. This score reflected the production of different macrostructural sequences, such as attempt-outcome (AO), goal-attempt (GA), goal-outcome (GO) and goal-attempt-outcome (GAO). The number of points awarded for each sequence reflected the level of complexity involved: 0 points if no sequence was produced (e.g. O only or A only), 1 point for an AO sequence (the least complex), 2 points for a GA or GO sequence, and 3 points for a full GAO sequence (the most complex). The scores obtained for each episode were then added, for a total score of 9 points maximum. Second, using a similar scoring system, we identified the highest level of narrative complexity reached by a child in each language. The maximum score was 3 points (for GAO). Third, we calculated, for each group and for each language, the percentage of children attaining the different levels of complexity.

Internal state terms

In the current study, use of ISTs was examined in each story told in LB and in French. In accordance with MAIN guidelines, ISTs were classified into different categories, including perceptual terms, physiological state terms, emotion terms, mental verbs, and linguistic verbs (see below for examples). For each language, we calculated the number of ISTs produced in each category and the total number of ISTs produced. For this measure, no maximum number of ISTs was expected. Finally, since most of the time, the ISTs retained for analysis were provided only once by the children, we did not conduct any type-token analysis.

Note that for all the measures described above, code-switching was not penalized. One point was attributed to each component when produced by the child regardless of whether code-switching had occurred or not. When it was used, code-switching was restricted to particular lexical items, such as *fox* and *crow*. There were no instances of code-switched ISTs. Finally, code-switching didn't concern alternate use of sentences in the two languages.

5. Results

Since most data followed a nonnormal distribution, either within each age group or for the whole sample, nonparametric tests were used for all statistical analyses, with Dunn-Bonferroni corrections implemented for multigroup comparisons. Effect sizes were calculated throughout, including Kendall concordance coefficients for Friedman repeated measure tests.

5.1 Participants and descriptive statistics

The characteristics of the participants regarding conceptual vocabulary skills, exposure to narratives, and language dominance are reported in Table 1.

Table 1. Participants: Age, conceptual vocabulary accuracy rates, exposure to narratives index (ENI) for LB and French, and language dominance index (LDI) (Mean (SD) and range)

	G4–5 ($n=15$)	G6–7 ($n=32$)	G8–9 ($n=22$)
Age	4.9 (0.3) 4.3–5.3	6.9 (0.6) 6.1–8.2	9.0 (0.6) 8.0–9.9
Conceptual vocabulary	62.6 (4.3) 51.9–68.5	79.1 (4.2) 67.3–87.1	83.5 (2.3) 80–87.1
ENI-French	2.5 (1.4) 0–4	3.1 (1.8)[a] 0–6	3.8 (1.3)[b] 0–6
ENI-LB	2.1 (1.3) 0–5	1.5 (1.4)[a] 0–5	2.5 (1.5)[b] 0–5
ENI-Global	4.6 (1.7) 2–7	4.6 (2.6)[a] 0–10	6.3 (2.4)[b] 1–10
LDI	13.1 (14.6) −26.8–+27	15.3 (14.5)[a] −17.8–+40	22.1 (13.5)[b] −4.3–+41

a. Based on 30 children

b. Based on 21 children; the ENI in each language ranged from 0 (least exposed to narratives) to 6 (most exposed to narratives); the global ENI ranged from 0 to 12; the language dominance index ranged from −49/50 (least dominant in LB) to +49/50 (most dominant in LB), the maximum and minimum scores depending on the number of contexts of language exposure before age 4.

Regarding conceptual vocabulary, all children had normal scores on the task. A Kruskal-Wallis nonparametric test revealed a significant difference across the three age groups ($X^2(2,69)=45{,}004$, $p<.001$, $E^2_R=.662$). Pairwise comparisons yielded systematic significant differences between groups ($p<.01$), meaning that performance on the conceptual vocabulary task significantly increased with age.

With respect to exposure to narratives, Table 1 shows that the ENI in French (ENI-FR) tended to be higher than the ENI in LB (ENI-LB), significantly so in G6–7 ($Z(32)=-3.698$, $p<.001$, $r=.462$) and G8–9 ($Z(22)=-3.027$, $p=.002$, $r=.456$), but not in G4–5. There was a significant age effect for both ENI-FR ($X^2(2,66)=6.331$, $p=.042$, $E^2_R=.097$) and ENI-LB ($X^2(2,66)=6.202$, $p=.045$, $E^2_R=.095$), with pairwise comparisons revealing a significantly higher index for French in G8–9 compared to G4–5 ($p=.037$). There were no significant inter-group differences for LB. The global ENI increased significantly across the three age groups ($X^2(2,66)=7.043$, $p=.030$, $E^2_R=.108$), particularly between G6–7 and G8–9 ($p=.037$).[5]

Finally, information from the PaBiQ revealed that most children had a high positive LDI (> 10), indicating that they were dominant in LB: 60% of the children

5. There was no significant increase between G4–5 and G8–9 despite similar mean global ENIs in G4–5 and G6–7, which may have to do with the distribution of the data.

in G4–5 (9/15), 62% in G6–7 (18/29), and 77% in G8–9 (17/22). Only one child in G4–5 and one in G6–7 had a high negative LDI (–26.7 and –17.8 respectively), indicating French dominance. The rest of the children had low LDI, suggesting more balanced bilingualism (see Appendix B). There was no significant difference across the three age groups for LDI ($X^2(2,66) = 4.397$, $p = .111$).

5.2 Story structure

The story structure scores obtained for each group and for each language are reported in Figure 1. Significant differences across the three age groups were found for both LB ($X^2(2,69) = 32.696$, $p < .001$, $E^2_R = .481$) and French ($X^2(2,69) = 37.855$, $p < .001$, $E^2_R = .557$). Pairwise comparisons revealed significantly lower performance in G4–5 compared to each of the other groups, for each language ($p < .01$ throughout). No significant differences were found between G6–7 and G8–9 in either language despite a higher mean in G8–9 than in G6–7 (LB: 8 vs. 7.1; French: 7.9 vs. 6.6).

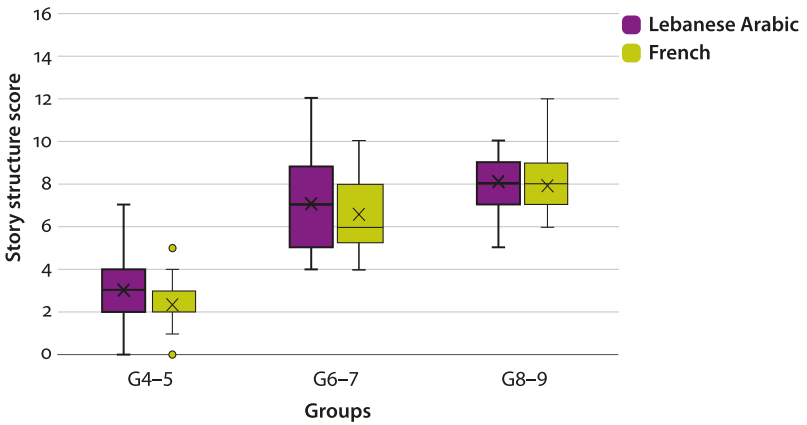


Figure 1. Story structure scores (max. 17) in each age group and in each language

Comparisons between the scores in LB and French revealed no significant language effect in any of the three age groups. The scores obtained in LB and in French were also strongly correlated in the entire population sample ($r_s = .786$, $p < .001$).

A detailed account of the production of the different episode components is presented in Table 2.

A significant age effect was found across the three groups, for most episode components (the same ones for each language), i.e. initiating events, goals,

Table 2. Mean percentage (SD) of overtly produced episode components by type in each group and in each language, and Kruskal-Wallis intergroup comparisons

Language	Episode component	G4–5 ($n=15$)	G6–7 ($n=32$)	G8–9 ($n=22$)	p value	Effect size
LB	Setting	3.3 (12.9)	17.2 (27.3)	20.5 (25.2)	.080	-
	Initiating event	8.9 (15.3)	39.6 (33.3)	40.9 (32.4)	.003	.175
	Goal	13.3 (21.1)	38.5 (22.6)	39.4 (31.9)	.004	.163
	Attempt	44.4 (27.2)	82.3 (23.9)	92.4 (17.6)	< .001	.369
	Outcome	28.9 (27.8)	58.3 (21.7)	74.2 (20.4)	< .001	.267
	Reaction	2.2 (8.6)	6.3 (13.2)	9.1 (18.3)	.420	-
French	Setting	10.0 (20.7)	17.2 (27.3)	22.7 (25.5)	.282	-
	Initiating event	8.9 (15.3)	44.8 (26.2)	34.8 (33.3)	< .001	.255
	Goal	13.3 (16.9)	36.5 (24.5)	40.9 (34)	.008	.142
	Attempt	15.6 (21.3)	67.7 (32.2)	89.4 (21.5)	< .001	.482
	Outcome	26.7 (31.4)	54.2 (25)	71.2 (25.8)	< .001	.256
	Reaction	6.7 (13.8)	6.3 (13.2)	13.6 (19.7)	.273	-

attempts and outcomes. In all these cases, pairwise comparisons revealed that the performance of the G4–5 group was significantly lower than that of the two other groups, for each episode component, and for each language. No significant differences were found between the G6–7 and G8–9 groups on any measure. Finally, no significant age effects were observed for settings and reactions, for which performance remained low throughout, including in the older group of children.

No language effect was found for any age group for any episode component, except for attempts in G4–5 and G6–7 where performance in LB was significantly higher than in French (G4–5: $Z(15)=-2.511$, $p=.012$, $r=.458$; G6–7: $Z(32)=16.5$, $p=.010$, $r=.321$). In G4–5 performance was particularly different in the two languages regarding the second episode, with 9/15 children expressing the attempt in LB versus 1/15 in French.

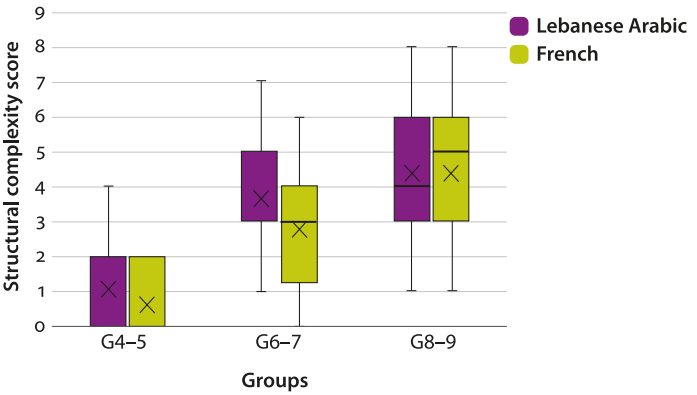
Comparing performance on goals, attempts and outcomes, Friedman tests showed significant differences across the three episode components in each age group and for each language (except for French in the G4–5 group where no significant differences were found).[6] Pairwise comparisons showed that goals were significantly less frequently used than attempts throughout, with no significant

6. For LB, Friedman analyses yielded the following results: G4–5: $Q(2,15)=8.714$, $p=.013$, $W=.290$; G6–7: $Q(2,32)=22.950$, $p<.001$, $W=.359$; G8–9: $Q(2,22)=21.028$, $p<.001$, $W=.478$. For French, the results were the following: G6–7: $Q(2,22)=15.571$, $p=.013$, $W=.243$; G8–9: $Q(2,32)=14.282$, $p=.001$, $W=.325$.

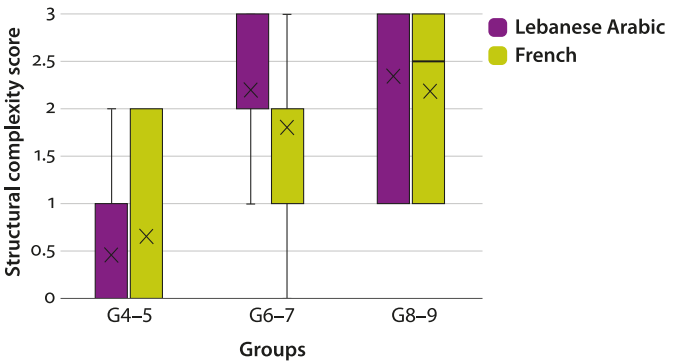
differences occurring between attempts and outcomes which tended to be the most frequently used story components.

5.3 Story structure complexity

Group results on complexity scores are represented in Figure 2, based on two methods of calculation: total complexity across the story (maximum 9 points) and most complex episode (maximum 3 points).



a. Total complexity (max. 9)



b. Most complex episode (max. 3)

Figure 2. Complexity structure scores in each age group and in each language based on two calculation methods (horizontal line = median, cross = mean)

The same tendencies were observed for each calculation method. First, as was the case for story structure, a significant age effect was found for story

structure complexity scores. Significant differences arose across the three age groups for LB (total complexity: $X^2(2,69)=25.325$, $p<.001$, $E^2_R=.372$; most complex episode: $X^2(2,69)=28.499$, $p<.001$, $E^2_R=.419$) and French (total complexity: $X^2(2,69)=29.035$, $p<.001$, $E^2_R=.427$; most complex episode: $X^2(2,69)=18.107$, $p<.001$, $E^2_R=.266$). Pairwise comparisons showed that in both languages, performance in the younger group was significantly lower than in each of the two other groups ($p<.01$ throughout), with no significant difference arising between G6–7 and G8–9.

Comparing performance in the two languages revealed no language effect for G4–5 and G8–9, but a significantly higher score for LB than for French was found in G6–7 (total complexity: $Z(32)=-2.262$, $p=.024$, $r=.283$; most complex episode: $Z(32)=-2.128$, $p=.033$, $r=.266$). Nevertheless, the scores on structural complexity calculated for each language were strongly correlated in our sample, regardless of the way structural complexity was computed (total complexity: $r_s=.682$, $p<.001$; most complex episode: $r_s=.614$, $p<.001$).

Finally, for each level of narrative complexity we looked at the proportion of children concerned, within each group, and for each language (Figure 3).

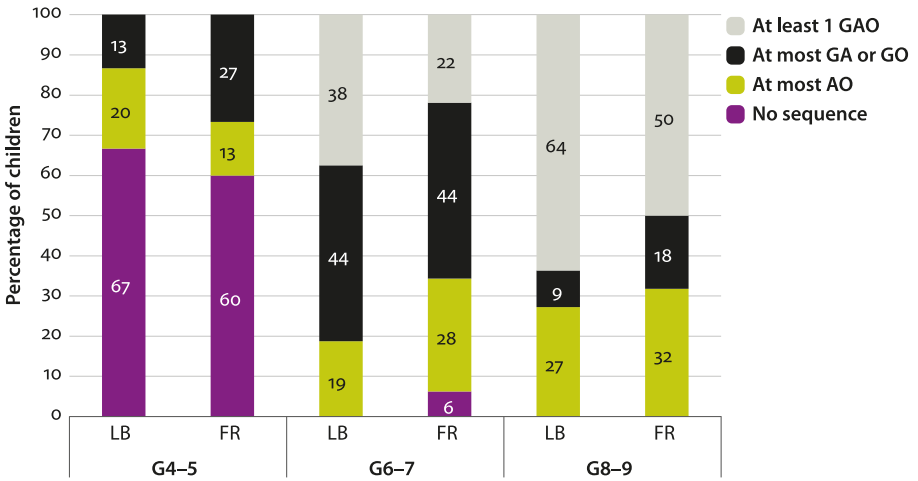


Figure 3. Percentage of children producing narrative sequences of different levels of complexity in each group and in both LB and French (FR)

The least complex level of narrative complexity (attempt or outcome only) was largely restricted to the youngest children (most of them producing no narrative sequences). No significant differences were observed between the proportion of children who used at most AO sequences, the next level of complexity, across the different groups, in any language. Concerning the percentage of children

using at most one GA or GO sequence, which is more complex than AO, there was an increase between G4–5 and G6–7 in both LB and French which didn't reach significance in either language. Between G6–7 and G8–9 a decrease was observed for these sequences, which was significant for LB ($X^2 = 7.511$, $p = .006$) and almost significant for French ($X^2 = 3.567$, $p = .050$). Regarding GAO sequences (the most complex level), none was found in the youngest children. GAO sequences started occurring in G6–7, and even more frequently so in G8–9, with a notable increase in the percentage of children using at least one GAO sequence between the two groups, which was significant for French ($X^2 = 4.641$, $p = .031$) and nearly significant for LB ($X^2 = 3.567$, $p = .059$).

Comparing the distribution of children according to the complexity levels of their narratives in each language, exact Fisher tests revealed no language effect in any group (G4–5: $p = .762$; G6–7: $p = .302$; G8–9: $p = .703$). In particular, lack of narrative sequences was observed in G4–5 in similar proportions for each language (around 60% of the children), and the percentages of children in G8–9 who produced at least one GAO sequence in LB and in French were very close (64% and 50% respectively). With respect to the proportion of children who produced at most GA or GO in G6–7, it was identical for each language (44%).

5.4 Use of ISTs

The number of ISTs produced by the children was generally low, on average less than one in the youngest group, around two in G6–7, and about 2.5 in the older groups (Figure 4).

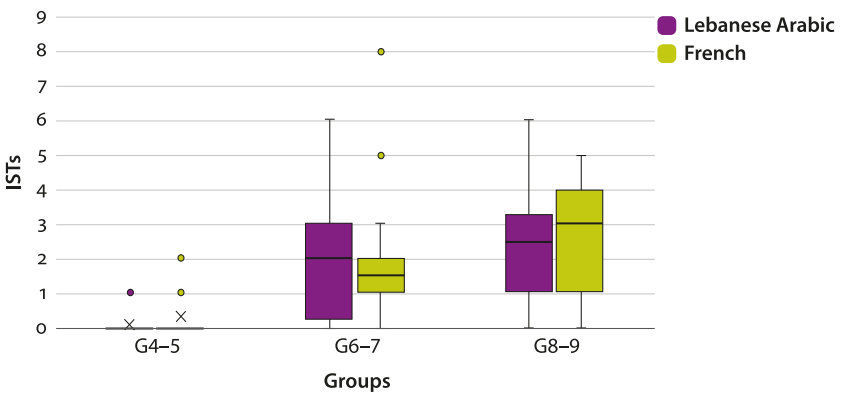


Figure 4. Number of internal state terms (ISTs) in each age group and in each language (horizontal line = median, cross = mean)

There was a significant difference between age groups for both LB ($X^2(2,69)=24.469$, $p<.001$, $E^2_R=.360$) and French ($X^2(2,69)=23.234$, $p<.001$, $E^2_R=.342$), with a significantly lower performance for the younger group compared to each of the two other groups ($p<.01$ throughout) and no difference between G6–7 and G8–9, in either language. No language effect was found in any age group, and a strong correlation was observed between the number of ISTs produced in LB and in French ($r_s=.517$, $p<.001$).

ISTs were broken into different categories (perceptual terms, physiological state terms, consciousness terms, emotion terms, mental verbs, and linguistic verbs). Figure 5 displays the mean number of each IST type used in each language in each age group. The majority of ISTs were perception terms (e.g. /ʃe:jif/ and *il regarde* 'he's looking'), followed by emotion terms (e.g. /mabsu:t/ 'he's happy' and *il a peur* 'he's scared') and mental state terms (e.g. /baddo/ and *il veut* 'he wants'). Only one consciousness term was found (/ʕajʃi:n/ 'they are alive').

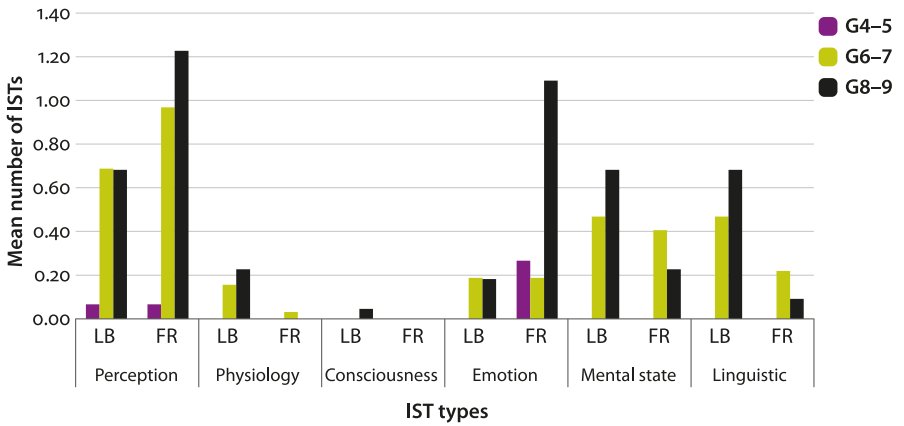


Figure 5. Mean number of IST types used in each age group in each language

There were few physiological state and consciousness terms used in any age group, in either language. Age effects were found for perception verbs for both LB ($X^2(2,69)=9.075$, $p=.011$, $E^2_R=.133$) and French ($X^2(2,69)=23.271$, $p<.001$, $E^2_R=.342$). An age effect was also found for emotion verbs, but only for French ($X^2(2,69)=16.411$, $p<.001$, $E^2_R=.241$). This was mainly due to frequent use of the verbal phrase *avoir peur* 'be afraid' in French, in contrast to LB. Another age effect was found, in LB only, for mental verbs ($X^2(2,69)=11.698$, $p=.011$, $E^2_R=.172$) and linguistic verbs ($X^2(2,69)=10.825$, $p=.004$, $E^2_R=.159$). When telling the stories in LB, children often used /baddo/ or /badda/ 'he/she wants' (e.g. /l-bsajne badda tʕarbiʃ ʕa ʃʃaʒra/ 'the cat wants to climb the tree') and they tended to make the

characters talk, reporting their speech with the verb /ʔa:l/ 'say'. These strategies were used to a lesser extent in French.

A language effect was found in G8–9 for emotion verbs (with higher use in French), mental verbs and linguistic verbs (with higher use in LB) ($p < .01$), for the reasons explained above. No language effect was found in the two younger groups for any ISTs. Note that on the whole verbs of perception were used throughout (especially in G6–7 and G8–9), with no language effect in any age group.

5.5 Impact of bilingualism factors and vocabulary

As can be seen in Table 3, all macrostructure production measures were significantly correlated with age and with conceptual vocabulary (for both languages). Regarding language dominance, as measured by the LDI, significant correlations were restricted to structural complexity scores in each language.[7] However, at the individual level, discrepant combinations between LDI and total complexity scores were found, namely children with a high LDI (suggesting dominance in LB) and high complexity scores in French, and children with a low LDI (suggesting dominance in French) and high complexity scores in LB. For example, in G8–9, a child had an LDI of 25 and a complexity score of 7 (out of 9) for French (and 5 for LB). Conversely, a child in G6–7 had an LDI of –17.8 and a complexity score of 6 in LB (and 4 in French). Finally, no significant correlations were found with exposure to narratives, in LB or in French. Note that with respect to ISTs, no significant correlations emerged involving exposure to narratives and language dominance within each IST category.

In order to determine which variables best predicted performance on macrostructure, we ran several forward stepwise multiple regression analyses with story structure scores, structural complexity scores and number of ISTs as dependent variables, for each language. In each analysis, we introduced age and conceptual vocabulary as independent measures. For structural complexity, LDI was added as an additional independent measure, following the results of the correlation analyses (Table 3). Results show that conceptual vocabulary was almost always the first variable to appear in the regression model, in a significant regression equation, and that age was almost systematically excluded (see Appendix C). For story structure, conceptual vocabulary explained 53.6% of the variance for LB and 54.9% for French. Regarding structural complexity, conceptual vocabulary accounted for 42% of the variance for LB and 28.7% for French (with LDI con-

7. Note that correlation analyses involving structural complexity were based on cumulative complexity scores, as their range (0–9) was wider than the scores based on the most complex episode (0–3).

Table 3. Bivariate correlations between story structure scores, structural complexity scores and number of internal state terms (ISTs), and age, exposure to narrative index (ENI), language dominance index (LDI) and conceptual vocabulary

	Age	ENI-LB	ENI-FR	ENI-Global	LDI[a]	Vocabulary
Story structure (LB)	.624**	.055	.207	.172	.15	.613**
Story structure (FR)	.653**	.063	.172	.170	.104	.635**
Structural Complexity (LB)	.533**	−.225	.066	.066	.263*	.578**
Structural Complexity (FR)	.529**	.019	.019	.019	.317**	.513**
ISTs (LB)	.530**	.093	.020	.020	.012	.532**
ISTs (FR)	.547**	−.049	.057	.057	.198	.532**

a. $n = 63$; LB = Lebanese Arabic; FR = French * $p < .05$ ** $p < .01$

tributing an additional 6.2%). As to ISTs, different tendencies were found between the two languages. While conceptual vocabulary explained 24.3% of the variance in French (with age being excluded), age was the retained factor for LB, accounting for 28.2% of the variance.

Finally, in order to further investigate the impact of exposure to narratives, language dominance and conceptual vocabulary on macrostructural skills, we compared, within G6–7 and G8–9, and for each language, the children who produced at least one GAO episode with those who produced episodes of low level of complexity, i.e. no narrative sequence (see Appendix D). Although we acknowledge that this analysis was based on small subgroups of children in some cases, Mann-Whitney nonparametric tests revealed almost no significant differences between the subgroups on any variable. Notably, children who produced at least one GAO episode in a particular language didn't tend to be more dominant in that language, nor had they received more exposure to stories in that language or in general.

6. Discussion and conclusion

In this paper, macrostructural skills of 69 bilingual LB-French children aged 4 to 9 were investigated using MAIN (Gagarina et al., 2012). Two stories were generated in each of the children's language, Baby Birds for LB and Baby Goats for French. Our research investigated potential age and language effects on production of story structure components (story structure complexity, and ISTs) and the potential relationship between these measures of macrostructural abilities and language dominance, exposure to narratives, and lexical skills.

6.1 Research question 1: Story structure

Results on the presence of main episode components in the children's narratives revealed a significant age effect in each language, with children aged 4–5 performing significantly worse than children aged 6–7 and children aged 8–9. In addition, language effects were very limited, meaning that performance was largely similar in both languages, which is consistent with findings reported in other MAIN-based studies (Bohnacker, 2016; Gagarina, 2016; Kunnari et al, 2016). A particularity of our study is that it involved children that were older than the children investigated in other MAIN studies, which mainly focus on ages 4 to 7 (but see Gagarina, 2016), which allowed us to confirm the similarity in performance across the children's languages that was observed at younger ages.

In other MAIN studies of macrostructure production, while a significant age effect was reported, performance seemed to differ across the child's languages, favoring either the majority or the minority language (Lindgren, 2018; Öztekin, 2019; Roch et al., 2016). In contrast to the bilingualism contexts of those studies, the children in our study were acquiring two majority languages, meaning that they had sustained exposure to both languages. In particular, the children were all attending private schools in Lebanon, where the principal medium of teaching is French, accounting for about three quarters of teaching hours (Hoyek, 2004). This situation is similar to Kunnari et al.'s (2016) study in which the two languages under investigation, Swedish and Finnish, are national languages in Finland. It thus appears that early and sustained exposure to literacy plays an important role in the development of macrostructural skills. As literacy skills develop, exposure to written narratives increases, which has been shown to play a role in comprehension of macrostructure (Dickinson & Smith, 1994). This could in turn have a positive effect on macrostructure production. Children in Lebanon start learning written language as of 5 years of age, which could account for the significant improvement in macrostructure performance in our children between ages 4–5 and 6–7. It is unclear at this point what factors or combinations of factors could account for the lower performance in one of the child's two languages reported in several studies of macrostructure production via MAIN, which is further complicated by the variety of bilingualism contexts involved. Likely candidates in need of further investigation include literacy skills in the minority language, types and frequency of storytelling activities, and SES.

Focusing on individual episode components, performance remained low on settings and reactions in our three groups, whereas an age effect was found for initiating events, goals, attempts and outcomes in both languages, with highest performance being observed for attempts and outcomes. The latter result was also found in Lindgren (2018), Bohnacker (2016), and Öztekin (2019), and in

Gagarina, Bohnacker and Lindgren (2019) which investigated (monolingual) adults. The varying performance across the different episode components could be related to the saliency of the actions displayed in the pictures. Attempts and outcomes are clearly depicted in both Baby Birds and Baby Goats, and could thus be described more easily than other episode components such as goals and reactions, which needed to be inferred from the pictures. The ability to understand the goal of the protagonists is indeed central to narrative knowledge and to the development of macrostructural skills (see below) (Westby, 2012).

Few language effects were observed in the production of individual episode components, which confirms previous findings on the same stories (Lindgren, 2018). The only significant differences between the two languages concerned performance on attempts, with higher performance in LB than in French, which could be related to differences between the stories. In Baby Goats, which was told in French, the outcome of the second episode (*the fox caught the baby goat*) could easily be expressed without mentioning the attempt (via the verb *try*, for instance). Indeed, mentioning both the attempt and the outcome could result in odd statements (*the fox tried to catch the baby goat and he caught it*). In contrast, in Baby Birds the attempt (*the cat climbed up the tree*) and the outcome (*the cat caught the baby bird*) could more naturally appear together (*the cat climbed the tree and caught the bird*).

6.2 Research question 1: Story structure complexity

Intergroup comparisons on story structure complexity scores based on two different scoring schemes (cumulative complexity and most complex episode) yielded similar results: a significant age effect was observed, confirming results in other studies (Bohnacker, 2016; Lindgren, 2018). A significant language effect was found in G6–7 only. No language effect was reported for story structure complexity in other MAIN studies on bilingual children (Kunnari et al., 2016; Lindgren, 2018; Öztekin, 2019). However, when adopting a third way of capturing story complexity, i.e. the proportion of children who reached the different levels of episode complexity for each language, no statistically significant differences between LB and French were found in any group, including G6–7. This suggests that one has to be careful about interpreting results from story structure complexity, especially when based on only one measure (see also Lindgren, 2018, where the results differed according to how story complexity was measured). This notwithstanding, looking at the distribution of children using different levels of narrative complexity across different ages yielded interesting results. Clear age effects emerged, with the lowest level of complexity (no production of any narrative sequence) being largely restricted to the youngest group of children, and the highest level (pro-

duction of at least one GAO sequence) emerging in G6–7 and increasing significantly (almost so in the case of LB) between G6–7 and G8–9. This latter result, coupled with the increase between G4–5 and G6–7 in the proportion of children producing two-item sequences including a goal, points to the fundamental role of goal in the development of macrostructure abilities. Perhaps more surprising was that the proportion of children producing (simple) AO sequences remained stable throughout the groups, including in G8–9 (see below).

How to better account for story structure complexity development? Many studies on macrostructure production have used the appearance of at least one GAO sequence as a milestone of development, including in monolinguals (Gagarina et al., 2019). In our study, tracking the percentage of children using GAO at least once also yielded a clear developmental trend. Taking into account cumulative complexity is interesting because of its scoring range (0–9 points), which allows for more statistical analyses than a three-point-range measure (most complex episode). However, it is more sensitive to inter-episode variability (in terms of how easy inferences can be made from the pictures) than focusing on the percentage of children reaching a particular level. In turn, cumulative complexity scores may be more suitable for standardization across a wider age span. This would need to be investigated in bilinguals older than 10 years.

6.3 Research question 1: ISTs

Following the general findings for story structure and story structure complexity, an age effect and no language effects were found concerning use of ISTs, when taking into account the total number of items being produced. The age effect is akin to findings in monolingual children for whom a strong developmental trend has been identified over the entire childhood (5 to 12) (Ukrainetz et al., 2005). Although absence of language effects has been reported elsewhere (e.g., Kunnari et al., 2016), some differences between LB and French were encountered when breaking down ISTs into different categories. Emotion terms tended to be used more frequently in French, whereas mental verbs and linguistic verbs were more frequent in LB. These may be due to a story effect and to differences in the way stories are told in the two languages in general. In both stories, use of emotion terms related to fear was expected, but in Baby Goats, which was told in French, the state of fear (of a baby goat) is visually more salient than in Baby Birds (which was told in LB), where a baby bird is meant to be afraid of a cat. Preliminary results from an ongoing study counterbalancing the stories across the two languages suggest that this explanation is on the right track, as fear is more frequently expressed in LB (Baby Goats) than in French (Baby Birds). Current analysis of use of linguistic verbs in LB should also help us determine whether or not chil-

dren, when telling stories in LB, tend to use direct discourse more often than in French. Similarly, frequent use of the pseudo-verb */baddo/* and */badda/* '(s)he wants' in LB needs to be further investigated, in particular through comparisons with the forms that are used in French in similar contexts.

6.4 Research question 2: Effects of language dominance, exposure to narratives and language proficiency on macrostructure performance

Although nearly all the children in our study were simultaneous bilinguals, variation in language dominance was high. Interestingly, the impact of language dominance was limited. No significant correlations were observed with story structure and use of ISTs, and while LDI was significantly correlated with story complexity (based on cumulative complexity) in both French and LB, regression analyses showed that it was to be retained for French only (and with little contribution to the variance). This is comparable with what was obtained on comprehension of macrostructure by Fiani et al. (2020) on a subset ($n = 48$) of the children who took part in the present study.

Performance on production of macrostructure was not found to be significantly correlated with exposure to narratives either. This differed from what was reported for comprehension by Fiani et al., 2020, who found that exposure to narratives in French significantly correlated with comprehension of macrostructure in both French and LB. The lack of significant correlations in production could stem from the fact that the questionnaire only included information about exposure to stories, without taking into account the amount and nature of interactions between caregivers and children during storytelling. Interaction during reading activities would be important to measure as narrative skills have been shown to improve thanks to interactive reading (Lever & Sénéchal, 2011). Adding questions addressing such interactions (e.g., whether parents frequently ask questions to their child during storytelling) would be an interesting way of further exploring the impact of storytelling activities on macrostructure skills in bilingual children.

Interestingly, when focusing more specifically on those children who produced complex narratives (i.e., at least one GAO episode) versus those who produced less complex narratives (i.e., containing at most AO sequences), no significant differences were found with respect to language dominance and exposure to narratives. This, however, does not mean that language use and exposure did not play a role on performance on macrostructure in these children, as they were growing up in a context where both of their languages were part of their daily lives, notably at school, as seen above. Clearly, the relationship between bilingualism factors and development of macrostructural skills needs further exploration. For one, although French and LB are spoken in Lebanon, bilingualism profiles

differ across the country (Hafez, 2006). Such differences may explain the (limited) predictive power of language dominance (as measured by the LDI) for French that was observed in our data concerning story complexity, and possibly the lower performance in G6–7. A more detailed language use and exposure questionnaire, including language practices at school, could be designed in order to capture potential bilingualism effects on narrative abilities more precisely. The effects of bilingualism factors could also be further explored by investigating other language combinations that are part of the Lebanese context, such as Armenian-LB. In this situation Armenian is the home language, which is not a majority language and is related to a community with a strong desire to maintain its cultural and language identity, and LB, the societal language. This would also lead to interesting comparisons with the bilingual context of Turkish-Swedish children in Sweden described in Lindgren (2018) and Öztekin (2019).

Finally, all measures of macrostructure production (story structure, story structure complexity and ISTs) were significantly correlated with conceptual vocabulary, as measured via the standardized ELO-L test battery. Stepwise regression analyses almost systematically revealed lexical skills to be the most predictive factor of macrostructure production performance across all children. This echoes findings reported by Lindgren (2018), which involved children from a different context of bilingualism (with a home language and a majority language). Not surprisingly perhaps, this suggests that although macrostructural skills have been shown to develop as a function of children's abilities to understand the motivations of others, language abilities are still necessary to express those motivations, which translates into being able to express, for example, the goals of the protagonists of a story, the attempts for reaching these goals, the outcomes of these attempts and the protagonists' reactions concerning the outcomes. Interestingly, significant correlations between vocabulary and macrostructure skills have also been found in comprehension of macrostructure (Fiani et al., 2020), as comprehension questions necessitated use of specific terms in the answers provided by the children.

6.5 Limitations and future steps

One limitation of the current study is that the children all attended private schools and came from families with middle SES. It would be interesting to recruit children from families with lower SES, including children attending public schools, who represent about a third of school-aged children in Lebanon (31.7%) (Lebanese Pedagogical Research and Development Centre, 2018). This would allow us to investigate the predictive power of SES in the development of macrostructure skills. Another limitation is that some of our analyses, such as the

ones comparing children producing narrative episodes of high complexity and children with narrative episodes of low complexity, were based on a low number of subjects. Although this was addressed by using of nonparametric tests, it would be interesting in the future to recruit more children in order to get further insight as to the effect of bilingualism factors and vocabulary on macrostructural skills. Finally, the two languages targeted by our study (French and LB) could not be balanced across the two stories, meaning that there were potential confounds between language effects and story effects, although Baby Goats and Baby Birds were originally designed to follow identical plots, with a similar number of episodes of comparable complexity. This shortcoming is currently being addressed in an ongoing study involving children in which the stories are counterbalanced in the two languages. This should shed light, among other things, on use of different categories of ISTs in the two languages, and on the difference in story structure complexity between LB and French identified in G6–7. It would also provide more insight into the general issue of the comparability of the MAIN stories.

References

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

Bitetti, D., & Hammer, C.S. (2016). The home literacy environment and the English narrative development of Spanish-English bilingual children. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1159–1171. https://doi.org/10.1044/2016_JSLHR-L-15-0064

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 of Chinese-English bilinguals. *Bilingualism: Language and Cognition*, *14*(4), 570–578. https://doi.org/10.1017/S1366728910000362

Dickinson, D.K., & Smith, M.W. (1994). Long term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*, *29*(2), 105–120. https://doi.org/10.2307/747807

Fiani, R., Henry, G., & Prévost, P. (2020). Narrative comprehension in Lebanese Arabic-French bilingual children. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives* (pp. 31–60). John Benjamins. https://doi.org/10.1075/sibil.61.02fia

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

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Asssessment Instrument for Narratives. *ZAS Papers in Linguistics*, *56*, 1–156.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–269). Multilingual Matters. https://doi.org/10.21832/9781783093137-011

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

Hafez, S.-A. (2006). *Statuts, emplois, fonctions, rôles et représentations du français au Liban*. Paris: L'Harmattan.

Hoyek, S. (2004). Le français dans l'enseignement scolaire et universitaire au Liban. *Cahiers de l'Association Internationale des Études Françaises*, *56*, 49–56. https://doi.org/10.3406/caief.2004.1525

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

Leclerc, J. (2015). *Liban dans l'aménagement linguistique dans le monde*. Québec, CEFAN, Université Laval. http://www.axl.cefan.ulaval.ca/asie/liban.htm

Leseman, P.P.M., Scheele, A.F., Mayo, A.Y., & Messer, M.H. (2007). Home literacy as a special language environment to prepare children for school. *Zeitschrift für Erziehungswissenschaft*, *10*(3), 334–355. https://doi.org/10.1007/s11618-007-0040-9

Lever, R., & Sénéchal, M. (2011). Discussing stories: How a dialogic reading intervention improves kindergarteners' oral narrative construction. *Journal of Experimental Child Psychology*, *108*(1), 1–24. https://doi.org/10.1016/j.jecp.2010.07.002

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

Lindgren, J. (2019). Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language*, *39*(4), 412–432. https://doi.org/10.1177/0142723719844089

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

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

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

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, *37*(1), 49–67. https://doi.org/10.1017/S0142716415000417

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

Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics*, *5*, 375–393. https://doi.org/10.1146/annurev-linguistics-011817-045554

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: disentangling bilingualism from language impairment* (pp. 301–330). 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 Sciences*, *48*(6), 1363–1377. https://doi.org/10.1044/1092-4388(2005/095)

Westby, C. E. (2012). Assessing and remediating text comprehension problems. In Alan G. Kamhi & Hugh W. Catts (Eds.), *Language and reading disabilities* (pp. 163–225). Pearson/Allyn & Bacon.

Zebib, R., Henry, G., Khomsi, A., Messarra, C., & Hreich, E. (2017). *Batterie d'Évaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Liban Tests Editions.

Appendix A. Detailed review of MAIN-based studies of macrostructure production

Table 1. Mean story structure production scores (max 17 points), story complexity measures, and mean number of Internal State Terms (ISTs) reported in studies using MAIN (Baby Birds and Baby Goats stories, in tell mode)

						Story complexity		
Study	Type of bilingualism	Languages (and status)	Age range	*N*	Story structure	% children producing at least 1 GAO	% GAO out of all possible sequences	ISTs
Kunnari et al. (2016)	Simultaneous	Finnish (Maj)	5;0–6;6	16	4.6		6.3%	1.3
		Swedish (Min)	5;0–6;6	16	5.4		14.6%	1.7
Lindgren (2018)	Simultaneous	Swedish (Maj)	4;0–4;11	14	5.1		2.4%	
			5;0–5;11	16	6.2		6.3%	
			6;0–6;11	16	8.5		12.5%	
		German (Min)	4;0–4;11	14	4.3		4.8%	
			5;0–5;11	16	4.5		4.2%	
			6;0–6;11	16	7.1		10.4%	
	Simultaneous + Sequential	Swedish (Maj)	4;0–4;11	16	3.6		2.1%	
			5;0–5;11	16	4.2		6.3%	
			6;0–6;11	16	5.5		10.4%	
		Turkish (Min)	4;0–4;11	16	3.1		4.2%	
			5;0–5;11	16	5.3		6.3%	
			6;0–6;11	16	5.5		8.3%	
Bohnacker (2016)	Simultaneous + Sequential	Swedish (Maj)	4;11–5;11	19	5.7		7%[a]	
			6;0–7;9	33	7.1		7%	
		English (Min)	4;11–5;11	19	4.9		12%	
			6;0–7;9	33	7.0		12%	
Gagarina (2016)	Simultaneous + Sequential	German (Maj)	2;7–4;4	21	3.4	± 5%[b]		0.9
			6;5–7;5	15	8.5	± 50%		5.2
			7;11–10;6	22	9.7	± 40%		3.9
		Russian (Min)	2;7–4;4	21	3.5	± 5%		0.8
			6;5–7;5	15	7.5	± 40%		2.1

Study	Type of bilingualism	Languages (and status)	Age range	*N*	Story structure	Story complexity: % children producing at least 1 GAO	Story complexity: % GAO out of all possible sequences	ISTs
			7;11–10;6	22	6.9	± 65%		1.7
Öztekin (2019)	Simultaneous + Sequential	Swedish (Maj)	4;0–4;11	25	3.3		1%	
			5;0–5;11	23	4.3		± 6%[c]	
			6;0–6;11	26	5.6		± 8%	
			7;0–7;11	28	6.5		6%	
		Turkish (Min)	4;0–4;11	25	3.3		3%	
			5;0–5;11	23	4.8		± 5%	
			6;0–6;11	26	5.6		± 5%	
			7;0–7;11	28	6.3		13%	
Roch et al. (2016)	Sequential	Italian (Maj)	5–6	30	7.1	46.7%		2.8
			6–7;2	32	6.5	34.4%		3.4
		English (Min)	5–6	30	4.3	20.7%		1.7
			6–7;2	32	5.9	34.4%		2.7

Key: Maj = Majority language; Min = Minority language; GAO = Goal-Attempt-Outcome

a. The exact figures for each language were not reported in Bohnacker (2016). In this study the percentages provided at age 5 and 6–7 were said to be evenly distributed between the two languages.

b. The percentages were approximated from Gagarina's (2016) Figure 2 (p. 104).

c. The percentages for the 5- and 6- year-olds in Swedish and Turkish were approximated from Öztekin's (2019) Figures 5.14 and 5.15 (p. 180).

Appendix B. Language dominance index in the population sample

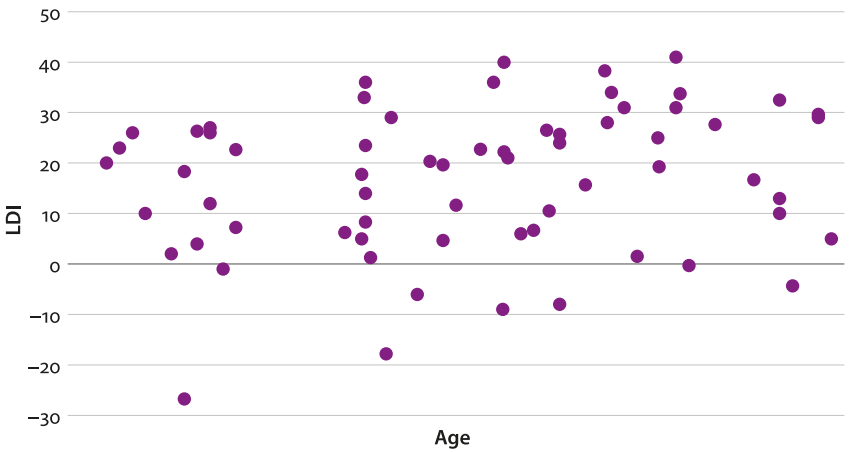


Figure 1. Distribution of individual Language Dominance Index (LDI) in our population

Appendix C. Stepwise multiple regression analyses

Table 2. Stepwise multiple regression: Story structure (LB) (dependent variable) and age and vocabulary (independent variables)

Variables entered[a]	R^2	ΔR^2	$R^2_{adj.}$	F(d.f.)	p	β	t	p
1 Vocabulary	.536	–	.529	77.434 (1, 67)	< .001	.732	8.800	< .001

a. Excluded variable in final model: Age

Table 3. Stepwise multiple regression: Story structure (FR) (dependent variable) and age and vocabulary (independent variables)

Variables entered[a]	R^2	ΔR^2	$R^2_{adj.}$	F(d.f.)	p	β	t	p
1 Vocabulary	.549	–	.543	81.710 (1, 67)	< .001	.741	9.039	< .001

a. Excluded variable in final model: Age

Table 4. Stepwise multiple regression: ISTs (LB) (dependent variable) and age and vocabulary (independent variables)

Variables entered[a]	R^2	ΔR^2	$R^2_{adj.}$	F(d.f.)	p	β	t	p
1 Age	.282	–	.271	26.276 (1, 67)	< .001	.531	5.126	< .001

a. Excluded variable in final model: Vocabulary

Table 5. Stepwise multiple regression: ISTs (FR) (dependent variable) and age and vocabulary (independent variables)

Variables entered[a]	R^2	ΔR^2	$R^2_{adj.}$	F(d.f.)	p	β	t	p
1 Vocabulary	.243	–	.232	21.521 (1, 67)	< .001	.493	4.639	< .001

a. Excluded variable in final model: Age

Table 6. Stepwise multiple regression: Structural Complexity (LB) (dependent variable) and age, vocabulary and LDI (independent variables)

Variables entered[a]	R^2	ΔR^2	$R^2_{adj.}$	F(d.f.)	p	β	t	p
1 Vocabulary	.420	–	.411	46.434 (1, 64)	< .001	.648	6.814	< .001

a. Excluded variables in final model: Age, LDI

Table 7. Stepwise multiple regression: Structural Complexity (FR) (dependent variable) and age, vocabulary and LDI (independent variables)

Variables entered[a]	R^2	ΔR^2	$R^2_{adj.}$	F(d.f.)	p	β	t	p
1 Vocabulary	.287	–	.276	25.757 (1, 64)	< .001	.536	5.075	< .001
2 LDI	.349	.062	.328	16.862 (2, 63)	< .001	.250	2.443	.017

a. Excluded variables in final model: Age

Appendix D.

Table 8. Age, exposure to narrative index (ENI), language dominance index (LDI) and conceptual vocabulary (from ELO-L) for children with narrative episodes of high complexity and children with narrative episodes of low complexity in G6–7, for each language

	Lebanese Arabic			French		
	High Complexity ($n=12$)	Low Complexity ($n=6$)	p value	High Complexity ($n=7$)	Low Complexity ($n=11$)	p value
Age	7.1 (.72)	6.8 (.6)	.437	6.8 (.64)	7.1 (.7)	.479
ENI-LB	.667 (.78)	2.2 (1.8)	.104	1.7 (1.2)	1.4 (1.5)	.479
ENI-FR	2.6 (1.8)	3.8 (2.3)	.279	1.9 (.9)	3.3 (2.0)	.151
ENI-Global	3.2 (1.8)	6.0 (3.7)	.279	3.6 (2.1)	4.6 (3.0)	.479
LDI	13.7 (18.3)	14.3 (11.9)	.879	21.0 (14.5)	12.0 (10.3)	.328
Vocabulary	80.5 (4.5)	75.5 (5.2)	.069	78.8 (3.5)	79.5 (4.9)	.860

Table 9. Age, exposure to narrative index (ENI), language dominance index (LDI) and conceptual vocabulary (from ELO-L) of children with narrative episodes of high complexity and children with narrative episodes of low complexity in G8–9, for each language

	Lebanese Arabic			French		
	High Complexity ($n=14$)	Low Complexity ($n=6$)	*p* value	High Complexity ($n=11$)	Low Complexity ($n=7$)	*p* value
Age	8.8 (.6)	9.2 (.7)	.239	8.6 (.5)	9.2 (.6)	.044
ENI-LB	2.2 (1.5)	3.0 (1.3)	.323	2.2 (1.3)	2.3 (1.5)	.887
ENI-FR	3.6 (1.5)	4.2 (1.3)	.467	3.9 (1.7)	3.7 (1.0)	.669
ENI-Global	5.8 (2.5)	7.2 (2.0)	.368	6.1 (2.5)	6.0 (1.7)	.813
LDI	25.9 (11.7)	15.5 (13.8)	.127	20.1 (15.4)	22.0 (15.2)	.962
Vocabulary	83.4 (2.7)	83.8 (1.9)	.779	83.7 (1.8)	84.4 (1.9)	.375

Address for correspondence

Rachel Fiani
Saint Joseph University of Beirut
Higher Institute of Speech and Langue Therapy (ISO)
Campus of Innovation and Sports
Damascus road
B.P. 17-5208 Mar Mikhaël
Beirut 1104 2020
Lebanon
rachelle.fiani@usj.edu.lb

Co-author information

Guillemette Henry
Saint Joseph University of Beirut
Higher Institute of Speech and Langue Therapy (ISO)
guillemette.henry@gmail.com

Philippe Prévost
UMR Inserm U1253 - iBrain
University of Tours
philippe.prevost@univ-tours.fr

Publication history

Date received: 15 May 2020
Date accepted: 8 July 2021
Published online: 7 September 2021