

# Language choice and code-mixing in a longitudinal study of an Estonian-English bilingual child

**Piret Baird**

School of Humanities  
Tallinn University  
Narva rd 25,  
Tallinn University  
E-mail: [pbaird@tlu.ee](mailto:pbaird@tlu.ee)

**Abstract.** The input that bilingual children receive influences their language proportions, language development, and code-mixing. Most studies on these topics have included early bilinguals whose input proportions undergo large changes in early childhood and whose parents use the one-parent-one-language family language policy. This paper examines the input-output proportions of an Estonian-English bilingual child over a period of 2.5 years (2;3-5;0<sup>1</sup>) using recorded spontaneous speech from a situation where the input language proportions did not change and where the family language policy was different from the one-parent-one-language policy that is presented in most studies: the family rotated the language they all spoke by the day of the week. Additionally, the child's code-mixing rate and her MLU scores are investigated to provide an overview of these factors in an unstudied input situation. Lastly, it is analyzed whether code-mixing by older siblings influences the code-mixing rate of the younger sibling. The results indicate that in the early phases of language development the child uses all the linguistic resources available to her, and as her language develops, she responds more in the language of the conversation and code-mixes less. However, there is also a period where the child unexpectedly almost stops speaking in Estonian regardless of the unchanged input. The data shows that code-mixed utterances are the longest, hence supporting previous research findings and indicating that code-mixing is a tool that helps the child communicate better. Code-mixing by siblings does not show any signs of affecting the younger sibling's code-mixing rate, though a more thorough analysis is necessary. Hence, the results indicate the importance of input and shed light on input effects in bilingual language acquisition in an understudied input situation.

**Keywords:** *early bilingualism, input, output, code-mixing, mean length of utterance*

## Dvikalbio estiškai ir angliškai kalbančio vaiko ilgalaikis tyrimas: kalbos pasirinkimas ir kodų kaita

**Santrauka.** Dvikalbių vaikų kalbų proporcijoms, raidai ir kodų kaitai daugiausia įtakos turi gaunama jų vartojamų kalbų įvestis. Iki šiol daugumoje šioms temoms skirtų tyrimų dalyvavo ankstyvieji dvikalbiai, kurių įvesties proporcijos ankstyvojoje vaikystėje stipriai keitėsi ir kurių tėvai taikė šeimos kalbų politiką, paremtą principu „vienas tėvas – viena kalba“. Šiame straipsnyje nagrinėjamos estiškai ir angliškai kalban-

<sup>1</sup> Age of the child is marked as years;months.

**Acknowledgements.** This research was partially funded by the Estonian Ministry of Education and Research grants EKKD33 and EKKD115. I would like to thank the anonymous reviewers for their valuable comments.

čio dvikalbio vaiko įvesties ir išvesties proporcijos per 2,5 metų laikotarpį (2;3–5;0) naudojant spontanineš kalbos įrašus. Tyrimo metu kalbų įvesties proporcijos nesikeitė, šeimos kalbų politika rėmėsi ne daugumoje tyrimų aptariamu principu „vienas tėvas – viena kalba“, o šeimos kalba buvo keičiama pagal savaitės dieną. Taip pat darbe tiriama vaiko kodų keitimo dažnis ir vidutinis pasakymo ilgis (angl. *Mean Length Utterance* – MLU) siekiant įvertinti šių veiksnių svarbą iki šiol netirtoje įvesties situacijoje. Galiausiai analizuojama, ar vyresniųjų brolių ir seserų kodų kaita turi įtakos jaunesniojo brolio ir sesers kodų kaitai. Rezultatai atskleidžia, kad ankstyvaisiais kalbos raidos etapais vaikas pasitelkia visus jam prieinamus kalbinius išteklius, o vystantis kalbai, jis dažniau atsako pagrindine pokalbio kalba ir mažiau keičia kodus. Tačiau esama ir tokio laikotarpio, kai netikėtai vaikas beveik nustoja kalbėti estiškai, nepaisant nepakitusios įvesties. Iš tyrimo taip pat matyti, kad pasakymai, kuriuose keičiami kodai, yra ilgiausi. Tai patvirtina ankstesnių tyrimų rezultatus bei rodo, kad kodų kaita yra priemonė, padedanti vaikui veiksmingiau bendrauti. Tyrime nepastebėta, jog brolių ir seserų kodų kaita turėtų įtakos jaunesniųjų brolių ar seserų kodų kaitos dažniui, tačiau tam patvirtinti būtina išsamesnė analizė. Tad, trumpai apibendrinant, rezultatai patvirtina įvesties svarbą ir atskleidžia jos poveikį dvikalbių vaikų kalbų įsisavinimui dar nepakankamai ištirtoje situacijoje.

**Raktažodžiai:** *ankstyvoji dvikalbystė, įvestis, išvestis, kodų kaita, vidutinis pasakymo ilgis.*

## 1. Introduction: input-output connections in (bilingual) children’s speech

For many decades, input was not researched much in terms of child language acquisition due to poverty of stimulus arguments. According to this assumption, input<sup>2</sup> was thought to be too poor for the child to be able to acquire language based on it, and instead children were thought to already be born with innate abilities to learn a language (Chomsky 1986). However, in recent decades the usage-based approach to language acquisition has emerged along with its emphasis on the importance of input (Bybee 2013; Langacker 1987; Tomasello 2003).

In the light of the usage-based theory, more research has been conducted on input-output effects in children’s language acquisition (see, for example, Ambridge et al. 2015; Behrens 2006; De Houwer 1990; Theakston & Lieven 2017). These studies have helped us understand the distributional information of child-parent speech, frequency effects in first language acquisition, the importance of multi-unit strings in language acquisition, and the role of input quantity and quality. In recent years, the study of input-output effects in the language acquisition of bilingual children has gained momentum. This is a necessary development as, due to refugee crises and globalization effects, more and more children are growing up bilingual.

Bilingual children’s input is divided between two languages, nevertheless yet their language development has been found to be similar to that of monolingual children (Hoff & Core 2015; Petitto et al. 2001). The rate of development of each of their languages is likely to be somewhat slower, and the languages involved do not necessarily develop at the same rate (Hoff & Core 2015). However, sufficient exposure to both languages is necessary for both languages to develop, as studies have indicated that the amount a language heard is connected to lexical and grammatical skills in that language (Hoff et al. 2012; Thordardottir 2011). For example, Thordardottir (2011) studied 5-year-old French-English simultaneous bilinguals’ language exposure and receptive and expressive vocabulary skills. She found

<sup>2</sup> In this study the term input is used rather than child-directed speech (CDS) because due to the nature of the material (interactions in a big family) it is quite difficult to differentiate input from CDS. It is not possible to always determine to whom a given turn was directed at and whether the modifications in the speech are due to the turn being directed at the child or is the way something is said in the family. The term input has also been preferred in other similar studies (see for example Quick et al. 2018, Quick et al. 2020).

that exposure to a language and performance in that language were related. Moreover, the language that receives more input has been found to develop more (Hoff et al. 2012; Pearson 2007; Place & Hoff 2011). For example, Hoff and colleagues (2012) found in their study of monolingual English and bilingual Spanish-English children aged 1;10-2;6 that bilingual children's vocabulary and grammar skills were related to the amount of input in that language.

Though on average bilinguals have less exposure to each of their languages, bilingual input does not necessarily always mean less input, because input amounts between children vary so greatly (De Houwer 2009; Hoff 2006). Thordardottir (2011) suggests that a balanced exposure is most likely to promote the successful acquisition of both languages. It has been found that balanced French-English bilingual 2-year-olds also have balanced vocabularies (David & Wei 2008). However, due to various life circumstances balanced input can be hard to achieve and maintain throughout a child's life. In most cases, the proportions of language input vary during childhood due to factors such as the start of daycare/school, extended visits from relatives, etc. Moreover, family language policy is a factor affecting input as, for example, with OPOL the child hears much more input from the parent who stays home with the child. Likely due to this, previous longitudinal studies have captured children whose input proportions of the two languages have undergone significant changes during the observation period, hence leaving a gap in our knowledge of language acquisition in the case of relatively stable input proportions.

Changes in the input levels of bilingual children are also reflected in their output language proportions. For example, Quick et al. (2021) studied 3 German-English bilingual children and their input-output relations. Each of the children's input situation was reflected in their output. When a child heard more German in his/her input, then her output also contained more German utterances. Furthermore, when one child's input proportions changed during the recording period, the change was reflected in her output language proportions: more English input was showcased by more output in English. In a similar study, Quick and colleagues (2020) analyzed the language proportions of four children with different language pairs, and again, each child's language input situation was reflected in their language output proportions.

In Western societies, where much of the research on language acquisition is carried out, the separation of languages in the input is emphasized (Gaskins et al. 2022). While parents are instructed to provide input to their children in both languages, they are often also instructed to keep the languages separate. One of the most common family language policy (FLP) methods that helps to follow this advice is one-parent-one-language (OPOL). In OPOL, each parent addresses the child in their native language. Research on bilingual children has used mostly families who practice OPOL family language policy, which means that we know less about bilingual language acquisition in situations where a different FLP is used. The current paper addresses this gap and uses data where the family employs a FLP where the languages are separated by days of the week: Estonian is spoken 3 days a week and English is spoken 4 days a week. Compared to OPOL, this provides a different and understudied input situation, and hence helps shed further light on the generalizability of the findings of previous research. In addition, the input proportions of the child (the proportion of input in Estonian and the proportion of input in English) were fairly balanced until the end of data collection and remained the same throughout the child's life. This relatively unchanged input proportion level is also a phenomenon not captured in other studies.

While parents are advised to keep the languages separate, young bilingual children are often reported to code-mix, which in this paper is understood as "all cases where lexical items and grammatical features of two languages appear in one sentence," meaning that the analyses in this article focus on intra-sentential code-mixing (Muysken 2000, 1).

## 1.1 Code-mixing in early bilingual language acquisition

Earlier explanations of child code-mixing saw it as confusion (Volterra & Taeschner 1978). It has now been established that bilingual children are not confused and are able to choose the right language according to the interlocutor (De Houwer 1990; Genesee et al. 1995; Lanza 1992). Nevertheless, most bilingual children code-mix, though to varying degrees. Studies of young bilingual children have reported code-mixing rates<sup>3</sup> from 0.5–49% (Gaskins et al. 2019; Genesee et al. 1995; Mishina-Mori 2011). It has been suggested that code-mixing is a natural phenomenon, and it could even be a developmental stage that bilingual children go through in their early language acquisition process (Byers-Heinlein & Lew-Williams 2013; Gaskins, Backus & Quick 2019; Mishina-Mori 2011).

The developmental nature of code-mixing suggests that children use mixing as a communicative strategy to employ all available linguistic resources in order to express themselves better. This seems to happen especially at the early stages of language acquisition. As noted above, language development of bilingual children does not mean that both languages develop at the same rate or that the same lexical and grammatical features of both languages are acquired simultaneously. Rather, depending on factors such as the frequency and salience of the construction in the input, certain lexical and grammatical constructions are likely to become more entrenched in one language than the other. As entrenched constructions are easier to activate (Schmid 2020), they are more likely to be used during speech, and for bilingual children this might mean code-mixing if both languages are used in one utterance. Support for this claim comes from studies that have looked at the mean length of utterance (MLU) of bilingual children and discovered that code-mixed utterances are longer and also more complex (Baird 2022; Quick et al. 2018, 2020, 2021). These results indicate that children use constructions that are more entrenched when they encounter a situation in which they would otherwise abort an utterance due to a lack of lexical or grammatical knowledge. Using a construction from the non-contextual language allows them to express themselves better (though code-mixing) and hence the code-mixed utterances are found to be longer (on average).

Code-mixing, being a developmental stage in bilingual language acquisition, is also supported by the fact that the amount of code-mixed speech decreases with age. Though there have been no longitudinal studies that investigate the code-mixing rate (e.g. how much of a child's recorded speech contains code-mixed utterances) of the same children from early childhood into teenage years, studies with young bilingual children report far higher code-mixing rates than studies with older bilingual children. For example, some of the highest code-mixing rates (above 40%) are found with 2-3-year-old bilinguals in studies by Mishina-Mori (2011), Baird (2022), and Gaskins et al. (2019). Redlinger and Park (1980), who analyzed the code-mixing rate of 4 children, found that the code-mixing rates of all children decreased with age. For example, Marcus's (2;0-2;5) rate of code-mixed utterances dropped from 30% to 21.2% and Henrik's (2;4-3;2) from 11.9% to 2.5%. Moreover, Virsu (2022) analyzed the number of code-mixed utterances in the speech of 4 sequential bilingual siblings and found that older siblings mixed less than the youngest 2-year-old. The 2-year-old's recorded speech contained 28% of code-mixed utterances, 5-year-old's 8.8%, 8-year-old's 1.4% and 10-year-old's 5.3%.

Many factors have been looked at to investigate what influences young bilinguals' code-mixing rate. These factors include mixing taking place due to lexical gaps (Nicoladis & Secco 2000), language dominance being a factor in mixing (Bernardini & Schlyter 2004; Genesee et al. 1995; Nicoladis & Genesee 1997), and caregiver's code-mixing influencing children's code-mixing rate (Comeau et al.

<sup>3</sup> Code-mixing rate in this work is defined as the amount or proportion of intra-sententially code-mixed utterances.

2003; Mishina-Mori 2011). Findings that connect caregiver's use of code-mixing in their own speech and children's code-mixing rate have given differing results. Comeau et al. (2003) studied six bilinguals (average age 2;4) and found in an experimental study that children patterned their language choices after the interlocutors and adjusted their code-mixing rates accordingly. Mishina-Mori (2011) studied longitudinally the language choices of two young Japanese-English bilinguals and found that the children did not always adjust their mixing rate to the mixing of their parents. Even though one of the participant's mother's mixing rate on average was 2.1%, her child mixed extensively when addressing her (average mixing rate of 48.9%). However, the same child did not mix as much when addressing his father, and the other participant of Mishina-Mori's study seemed to follow her parent's language choice patterns (though less so in her speech to her father in the first five monthly recording sessions). Moreover, Nicoladis and Genesee (1997) studied the speech of seven bilinguals from 2;0-3;6 and found that the parent's code-mixing rate correlated with that of their children at ages 3;0 and 3;6, but not at ages 2;0 and 2;6.

Besides parents, siblings are also an influential factor in bilingual families. However, not many studies have examined siblings' role in influencing each other to use code-mixing. Barron-Hauwaert (2011) has reported on survey data, but only on how many parents have observed their children using code-mixing when interacting with siblings at home. Jiménez-Gaspar and Arnaus Gil (2022) studied siblings' language choice and reported some findings on code-mixing, but the study lacked focus on code-mixing. Thus, it is still not clear whether and to what extent code-mixing by interlocutors, including siblings, affects the code-mixing rate of young bilingual children. Therefore, this study not only examines the rate of code-mixing longitudinally, but also attempts to investigate older siblings' influence on the younger sibling's code-mixing rate.

## 1.2 MLU as a measurement of language level

Mean length of utterance (MLU) is the most known and used tool to measure children's language development (Nieminen 2009). It was introduced in 1970s by Brown (1973) as a better indication of children's language development than age. MLU is calculated by taking the total number of words or morphemes and dividing it by the total number of utterances. In child language acquisition studies, MLU is usually either calculated in morphemes (MLU-m) or words (MLU-w). Studies of Estonian child language have used MLU-w because none of the child language corpora in Estonian are morphologically marked as there is no consensus on how to mark certain situations. Estonian has many fusional forms (for example, plural partitive) where it is difficult to determine the exact number of morphemes in a word.

Studies have used MLU for intra- and inter-individual comparison, but also for assessing language development cross-linguistically in the case of bilingual language acquisition (Quick et al. 2020). When doing it cross-linguistically, one needs to keep in mind the possibility of how the characteristics of different languages affect the results. For example, it is likely that for a given level of language development a bilingual child's MLU-w in Estonian will be lower than MLU-w in English because Estonian uses more case endings instead of prepositions (e.g. *with a doll* [3 morphemes] would be *nuku-ga* 'doll-COM' [2 morphemes] in Estonian).

Brown (1973: 54) also noted that once children reach MLU 4.5, the index has limitations. At this point the MLU depends more on the nature of the interaction than on the child's knowledge. It has been found as well that once children reach MLU 4.5, they are also able to increase the structural complexity of the utterance without increasing the length of the utterance (Chabon et al. 1982).



While MLU has been employed in many studies as a tool of children's language development, a few studies have specifically analyzed the connections between bilingual children's MLU's in each of their languages and their language input situations. For example, Quick and colleagues (2020) analyzed the speech of four bilingual children, each with a different language pair, and discovered that children's MLU scores reflected their input situations: the more a child received input in a given language, the longer and more complex his/her utterances in that language were. Moreover, they determined that code-mixed utterances were longer than monolingual utterances and suggested that entrenchment and activation play a role in helping bilingual children to achieve this greater communicative competence while code-mixing. Similar findings have been reported by Quick and colleagues (2018, 2021) for three German-English bilinguals (2;3-3;11).

### 1.3 Research questions

As can be seen from the above overview, many studies have been conducted to understand the role of input in bilingual language acquisition. However, it is also evident that most studies have used participants from OPOL families and the proportions of language input have changed during the recording period. Moreover, siblings as a source of input, including code-mixed input, have received little attention. Considering the above, the current study seeks to answer to the following research questions:

- (1) How do language proportions change during a child's language acquisition over a 2.5-year period (2;3-5;0) while keeping the input language proportions relatively unchanged?
- (2) How are the proportion of code-mixed utterances and mean length of utterance (MLU) related?
- (3) How does siblings' code-mixing rate affect the child's code-mixing rate?

## 2. Method

### 2.1 Participant and data

The focus participant of the study, Fiona, was an Estonian-English simultaneous bilingual child who was 2;3 years old at the beginning of the recording sessions. The mother is a native Estonian speaker and the father is a native English speaker, but both speak the other language at a high level. The family resides in Estonia and employs a family language policy where they speak Estonian 3 days a week and English 4 days a week. Occasional interactions with Estonians (the societal language) on English-speaking days meant that the language input of the child was fairly balanced. The child's exposure to media (cartoons, audio) was minimal and the exposure was in both languages. There was also input from playmates in both languages<sup>4</sup>, and a 2-week stay in an English-speaking country at age 4;10 when there was more input in English, though the family mostly followed their regular FLP, except when conversing with monolingual English speakers. The child started attending Estonian medium daycare at age 4;2, but attended 2 days a week on the days when the family spoke Estonian, so the balance of language input remained the same. The participant of the study has two older siblings who were 7;6 (Sister) and 5;3 (Brother) at the beginning of the first recording session. The family has followed the same family language policy since the birth of the first child and both siblings have adhered to this

<sup>4</sup> However, as a lot of the recordings were made during the COVID-19 pandemic, the input from playmates was less frequent than what it would be under normal circumstances.

family language policy during their upbringing. The parents rarely code-mix when speaking with the children, but the older siblings sometimes do, which is why their code-mixing was analyzed rather than that of the parents.

The data for this study involves three different datasets:

- 1) The first dataset was recorded from 2;3-2;11 (46h, 9364 utterances from Fiona, 2273 from Sister, 3806 from Brother). The recordings were made about once a week with one session being 1-1.5h long.
- 2) The second dataset was collected densely with recording sessions six times a week over a seven-week period (with a one-week gap in between) at the age of 3;1-3;2 (38h, 7126 utterances from Fiona).
- 3) The third dataset was recorded from 4;9-5;0 (11h, 2653 utterances from Fiona, 1408 from Brother). The recordings were made about once a week, and each recording session was about 1h long.

All recordings were made during play and meal times, with the mother and sometimes other family members present. The recordings were usually made by the mother, but sometimes also by the father.

## 2.2 Analyses

To conduct the analyses, each utterance was coded for language type: Estonian, English or mixed. Unintelligible utterances were excluded. Also, utterances with ambiguous language (e.g. *okay/okei*) were left out of the analysis. Thereafter, three analyses were carried out.

First, the proportion of Estonian, English and code-mixed utterances for each month was calculated to get an overview of Fiona's language proportions and their change over time (research question 1). Second, MLUs, in words, for Estonian, English and code-mixed utterances were calculated to see MLU changes over time and to analyze its relationship with the proportion of code-mixed utterances (research question 2). For code-mixed utterances only intra-sentential code-mixes were analyzed (inter-sentential code-mixes by the child can be seen in the language proportions as, for example, on Estonian days utterances in English are technically inter-sentential code-mixes because the other speakers answered in Estonian).

Third, recordings which included either or both of the older siblings were separated from the ones without siblings. Thereafter, the average rate of siblings' code-mixing was calculated for the first and third datasets. The second dataset was not included in this part of the analysis because almost all of the recordings were made with the siblings present. A paired t-test was performed on some of the data from the first dataset (data from Estonian days from ages 2;3-2;6 and English days from 2;6-2;11) to test whether the presence of siblings influenced the younger child's code-mixing rate (research question 3).<sup>5</sup>

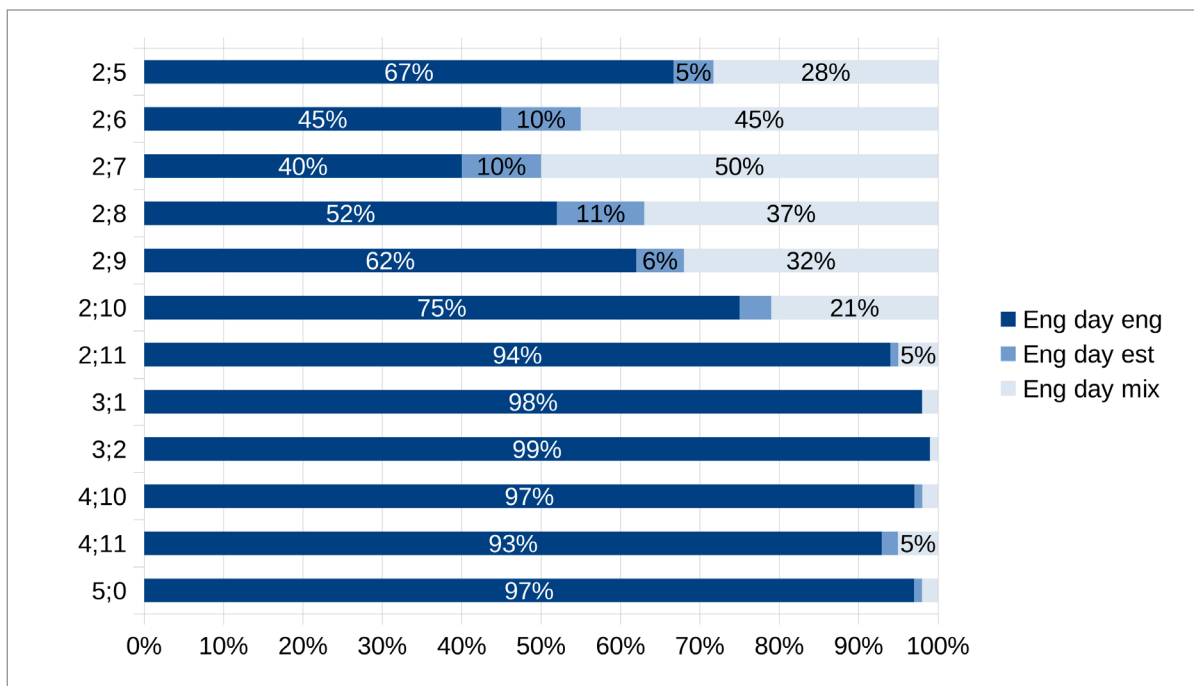
<sup>5</sup> As the siblings were not the main aim of the recording sessions, then their data was sporadic. In order to take into account the different recording settings (language spoken) and siblings presence vs non-presence only data from these months were used.

### 3. Results

#### 3.1 Language proportions

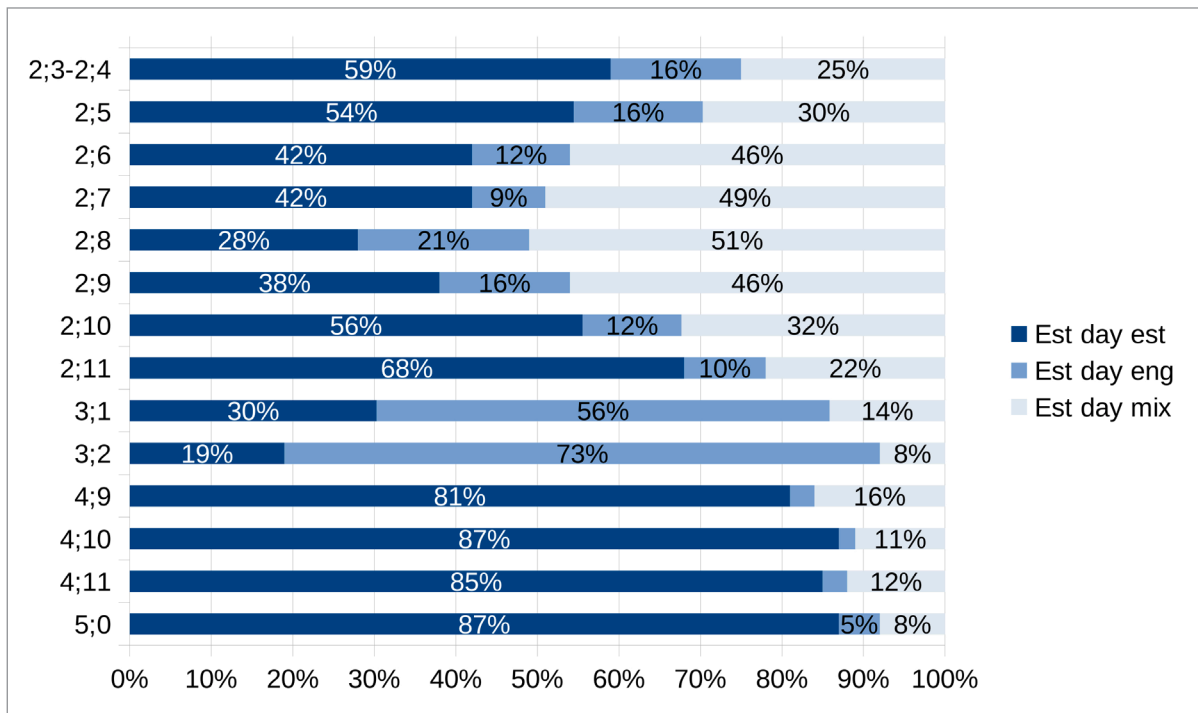
An analysis of the language proportions (research question 1) shows variation. Initially, the child uses a lot of code-mixed utterances, the use of which peaks around age 2;7-2;8, when about half of her utterances contain code-mixing (see Figure 1 and Figure 2). This is more than is usually reported in studies involving young simultaneous bilingual children (compare, for example, Quick et al. 2020, 2021). At the same time, the proportion of monolingual utterances follows the inverse trend. At age 2;3-2;5 over half to 1/3 of utterances are either monolingual Estonian or English, depending on the day the recording took place, after which the proportion of monolingual utterances decreases and starts to increase again when the proportion of code-mixed utterances starts to decrease. It is also notable that the code-mixing rate on Estonian days decreases later than on English days.

However, the numbers look very different at age 3;1 and 3;2. While the input has remained unchanged, the child strongly prefers to speak English. On English days almost all of her utterances are in English with very little code-mixing (1.4%-2.2%). On Estonian days she speaks more English than Estonian and uses code-mixed utterances at a higher rate than on English days. However, there is less code-mixing than at an earlier age (14% and 8%). While the child tends to speak English on Estonian days, the parents and siblings continue to respond in Estonian and occasionally remind the child to speak in Estonian.



**Figure 1.** Language proportions by age on English-speaking days





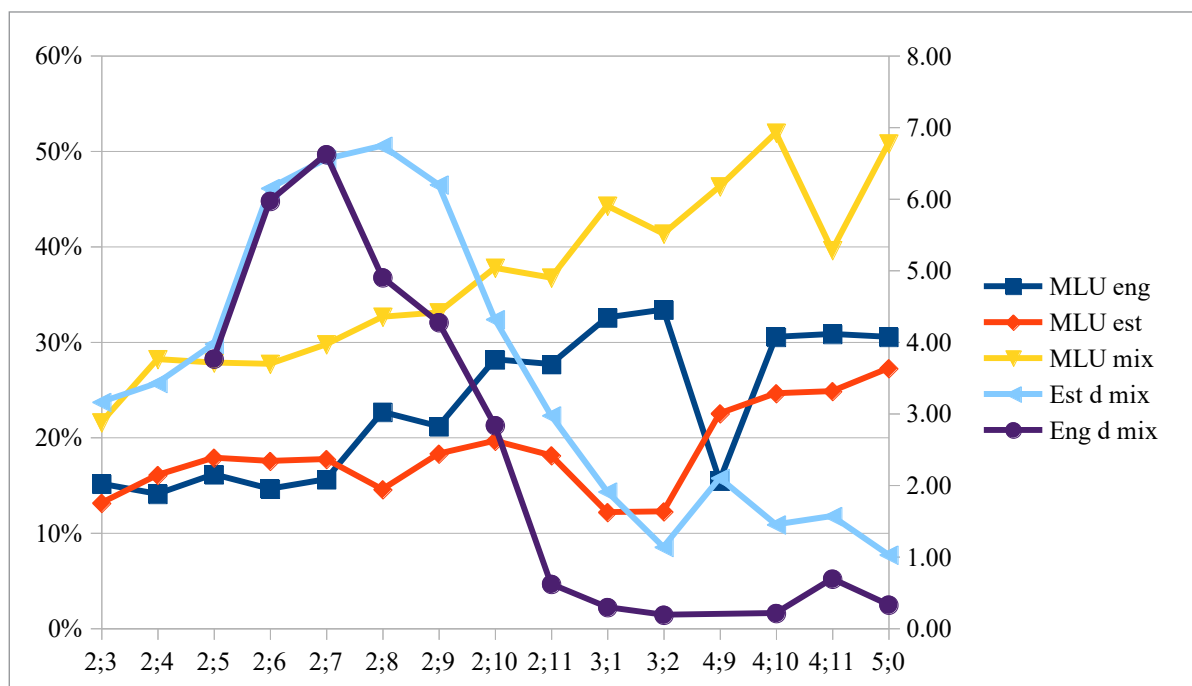
**Figure 2.** Language proportions by age on Estonian-speaking days

Without any changes in input proportions (though at age 4;2 she started attending Estonian medium daycare 2 days a week for 6-7h a day on days the family spoke Estonian), at age 4;9 the child speaks mostly in Estonian again on Estonian days (81%), almost no English (3%), but still uses code-mixing (16%). On English days the pattern has remained the same as at age 3;2: over 93% of utterances are in English with little Estonian (1-2%) and some code-mixing (2-5%).

### 3.2 Mean length of utterance scores

MLU in English starts rapidly increasing around age 2;7 when it goes from 2.08 to 3.69 by age 2;11 (see Figure 3). MLU in Estonian grows steadily from 1.75 (2;3) to 2.62 (2;10). Throughout the recording period code-mixed utterances have a noticeably higher MLU than monolingual utterances: from 2.87 at age 2;3 (example: *one kabi siin* ‘one pinecone here’) to 4.90 at age 2;11 (example: *I have only two rocks*’e ‘I have only two rocks-PTV.PL) to 6.77 at age 5;0 (example: *Kris vaata mul on selline lilla tongue* ‘Kris look I have such a purple tongue’).

In the dense dataset, the MLU for Estonian monolingual utterances has decreased to 1.62. However, almost half of those utterances are single-word responses ‘yes’ and ‘no’. Hence, this low MLU seems to not indicate her level of Estonian, but rather her unwillingness to speak much in Estonian. MLU in monolingual English utterances continued to increase and reached 4.45 (3;2). For code-mixed utterances, MLU went up to 5.51 (3;2) (example: *See girl läks with the stroller* ‘This girl went with the stroller’).



**Figure 3.** Development of MLU and percentage of code-mixed utterances in time. The MLU score on the right axis and percentage of mixed utterances on Estonian and English days on the left axis (Est d mix = percentage of code-mixed speech on Estonian days, Eng d mix = percentage of code-mixed speech on English days)

At age 4;9, MLU in Estonian is 3.0 and at age 5;0 3.63, while her MLU in English at the same period hovers around 4.0.<sup>6</sup> MLU for code-mixed utterances for this last recording period remained higher than MLU for monolingual utterances.

### 3.3 Siblings’ code-mixing rate as an influencing factor

To analyze whether siblings’ code-mixing influences the younger child’s code-mixing rate (research question 3), the average code-mixing rates of the two older siblings were calculated. The results show that the older sibling (Sister, whose data was only available in dataset 1) code-mixed less than in 4% of utterances in both languages. The Brother code-mixed more on Estonian days at age 5 than at age 8 (9% vs 2%) (see Table 1). At age 8, the Brother code-mixed about the same as the Sister at the same age (4%), but used more English on Estonian days (but not vice versa): 19% vs 0%. However, it should be noted that there were fewer utterances available for this time period.

**Table 1.** Siblings’ language proportions and code-mixing rates by language day

	Brother 5;4-5;10	Brother 7;10-8;0	Sister 7;7-8;1
Estonian day est	88%	77%	96%
Estonian day eng	3%	19%	1%
Estonian day mix	9%	4%	4%
English day eng	98%	99%	98%
English day est	0%	0%	0%
English day mix	2%	0%	2%

<sup>6</sup> At age 4;9 the graph shows an abnormally low MLU for English utterances, which is due to the fact that there were not enough English utterances on the recordings for that month (only 17) to capture the MLU correctly.

**Table 2.** T-test results for Estonian speaking data

	Without siblings	With siblings
Mean	29%	28%
Variance	0.16%	1.15%
Observations	6	3
t-statistic	0.24	
P-value (two-tail)	0.83	

**Table 3.** T-test results for English speaking data

	Without siblings	With siblings
Mean	27%	29%
Variance	4.77%	2.00%
Observations	4	5
t-statistic	-0.77	
P-value (two-tail)	0.49	

The results of the t-test show there is no difference in the code-mixing of the younger sibling whether or not the older siblings are present (see Table 2 and Table 3). The means for both, English and Estonian, data are close to one another and the variance is small. The t-test gives high p-values, which confirms that the difference in the means is insignificant.

## 4. Discussion

### 4.1 Code-mixing and MLU

The data shows that once the child reaches a certain MLU level in each language (around MLU 2.5), there is a steady decrease in code-mixing. This is especially evident for English utterances. Up until age 2;7 when her MLU for English utterances increased to 2.08, the proportion of code-mixed utterances also continued to increase from 28% to 50%. This indicates that as the child became more communicative, she maximized her resources and code-mixed when certain constructions were not available or entrenched enough in one language as has been suggested to happen by Quick et al. (2021). By age 2;8 her English MLU jumped to 3.02 and the amount of code-mixed utterances fell to 37% and continued to decrease to 1.4% by age 3;2 while her MLU of English utterances rose to 4.45.

A similar trend can be seen for Estonian. The rate of code-mixing on days when the family spoke Estonian is highest at age 2;8 (50.5%) when the MLU for Estonian is 1.94, but once it reaches 2.44 at age 2;9, the proportion of code-mixed utterances also starts to decrease. However, it should be noted that the MLU for Estonian does not continue to increase, but remains around 2.4-2.6 (and even drops at age 3,1-3;2, but more on this later), while the code-mixing rate continues to decrease. This is probably due to the fact that Estonian is a synthetic language and the child has reached a developmental stage where she is able to increase the complexity of her utterances by adding case and inflectional endings to words. Example (1) shows the difference between English and Estonian for the length of an utterance.

- (1) *Tahan mängida autoga liivakastis.*  
Want-PRS.1PL play-INF car-COM sandbox-INESS  
'I want to play with the car in the sandbox.'

Thus, while in example (1) 10 words are needed in English, 4 words are sufficient in Estonian; therefore, it is possible to express oneself in a syntactically more complex way without increasing much the length of the utterance.

MLU values were highest for code-mixed utterances continually throughout the recording period, as reported in previous studies (Baird 2022; Quick et al. 2018, 2020, 2021). These studies have suggested that this is due to entrenchment effects, as bilingual children use constructions that are more entrenched in the early stages of language acquisition due to easier activation. This leads to the use of both languages and subsequently to code-mixing, as in, for example, *kus su bellybutton on* 'where is your bellybutton' uttered by the child at age 4;9. Overall, these findings support the claims made by Gaskins, Backus and Quick (2019) that code-mixing is a developmental stage that early bilinguals go through. The need for code-mixing decreases as the language development in both languages reaches a level where the child is able to express herself well enough. However, some mixing remains because certain forms may still be more deeply entrenched in one language than the other, and possibly for other reasons.

Previous studies have found connections between input levels and language development in bilingual children, and this is also supported by the findings of this study. In their study of mono- and bilingual children's language development, Hoff and colleagues (2012) found that the development of bilingual children's vocabulary and grammar was related to the relative amount of input they received in that language. The fact that the MLU scores of the child in this study increased at the same rate reflects the largely balanced input she received. The somewhat higher MLU for English utterances could either be the result of slightly more exposure to English (4 days over 3 days, although the Estonian social language that surrounded the child whenever she left home could balance it out), or it could reflect the differences between the two languages being showcased in the MLU score. Studies have also shown (e.g. Quick et al. 2018, 2020, 2021) that input situations are reflected in children's MLU scores. Quick and colleagues (2020) found that for each child in their study, input proportions were associated with MLU scores. When a child received more input in German than in English, his MLU for German was higher than his MLU for English. In another study, Quick et al. (2021) found that the child whose input proportions of his two languages were fairly balanced also had her English and German MLU develop at the same pace.

## 4.2 Language proportions

The analysis showed that in the early stages of language acquisition the study participant generally followed the input language, as previous research has shown (e.g. Quick et al. 2021). However, the child used many code-mixed utterances and also sometimes used the language of the other day (e.g. English on days when the family spoke Estonian) when interacting with her family. This was especially so in the early phases of language acquisition. While many studies have found code-mixing to be present in early bilingual language acquisition, few have reported periods when about half of the child's utterances in both languages contain mixing. This raises the question of why some children mix so much more than others. One possible reason could be the environment in which bilingual children live. Young bilinguals have been shown to be sensitive to the language skills of their interlocutors, so it could be that bilingual children of parents who understand and speak both or all the languages in the

family also code-mix more. If a young child utters a word or phrase in one language and his/her parent does not understand it and indicates it repeatedly to the child, the child is likely to recognize it. On the other hand, if the child is understood regardless of the language, he/she is likely to recognize it, and if certain constructions are more entrenched and hence easier to activate, he/she is going to use them. However, while most studies of bilingual children do mention the language(s) spoken by a parent to the child, not all of them report whether a parent understands the other family language. This detail would be important to include in future studies of code-mixing in early bilinguals, to compare the mixing rates of children whose parents understand all the languages spoken by the child with those who do not, in order to gain further insight into children's mixing.

It has been noted in the literature that children's language output proportions follow their input trends (Quick et al. 2021) and that balanced exposure is most likely to guarantee the development of both languages (David & Wei 2008; Thordardottir 2011). This seemed to be the case for the participant in this study as well: balanced exposure in terms of quantity up until the age of 2;11 showed balanced output language proportions and similar MLU scores. However, the data showed a sharp decline in Estonian around age 3, which cannot be explained by a change in input. Example (2) shows an exchange between a parent and the child to exemplify the preference for English.

- (2) Mother: *Ma pean aknast ka vahepeal välja piiluma et mis need teised kaks sõpra teevad seal.*  
'I have to peek out of the window to see what the other two friends are doing there.'  
Child: *I want to look.*  
Mother: *Mängivad.* 'They are playing.'  
Mother: *Täitsa mängivad.* 'Totally playing.'  
Child: *But why Keia ei mängi?* 'But why Keia is not playing?'  
Mother: *Keia ka ju mängib.* 'Keia is also playing.'

If at age 2;11 68% of the utterances on Estonian days were uttered in Estonian, at age 3;2 it was only 19%, and the child also spoke on Estonian days mostly in English (73%). This happened even though the parents reported no changes in the input patterns. At age 3;0, there was a 2-week visit from an English-speaking aunt, during which the family spoke English when the aunt was present even on days when they usually would have spoken Estonian. This possibly could have had an influence on the child's unwillingness to speak Estonian. However, at the same time, the family spent about a week at age 3;1 visiting Estonian relatives, where again only Estonian was spoken when non-English speaking relatives were present, even on days when they would normally have conversed in English. This does not seem to have affected the proportion of Estonian, as the percentage of Estonian utterances continues to decline from age 3;1 to 3;2.

This indicates that there were other factors that played a role in this young bilingual child's language choice. Even though the language input proportions remained unchanged and balanced, at age 3 the child chose one language over the other. Hence, balanced input does not always guarantee balanced output. The same change in preference was not observed for the older siblings in any of the recordings, who kept following the family language policy of alternating languages on different days of the week. Input quality, including receiving input from a variety of native speakers, has been reported to be a factor influencing language development (Cameron-Faulkner & Noble 2013; Montag et al. 2015; Noble et al. 2019; Place & Hoff 2011). While no recordings were made between ages 3;3-4;9, parents report that the shift back to speaking both languages occurred gradually after age 4;2 when the child started attending part-time Estonian medium daycare, which only changed her input place, but not the language input proportions. This suggests that changing interlocutors (input quality) affected



language choice. This is reflected in her language output proportions at age 4;9 to 5;0, which again reflect the earlier balance, though there is more code-mixing on Estonian days than on English days.

This surprising and sudden preference for English could be due to individual preferences. In the same way that we may prefer one type of food over another in our daily lives, or choose our clothes according to our mood, bilingual children may prefer one language over another, and these preferences may change over time. Moreover, the language choice of a bilingual child does not seem to be always related to input patterns. Furthermore, the child may continue to prefer one language over the other even in a situation where parents and other interlocutors ask the child to use a certain language and model its use, as can be seen in example (3).

- (3) Mother: *Kuule aga kas sina minuga eesti keeles ei taha rääkida?* ‘Listen but don’t you want to speak to me in Estonian?’  
Child: *Mh.*  
Mother: *Teeme nii?* ‘Let’s do so?’  
Child: *Jah.* ‘Yes.’  
Mother: *Okei.* ‘Okay.’  
Mother: *Nii siis ma keedan meile riisi.* ‘so I will boil rice for us’  
Child: *And then this row.*

The individual difference is also evident in the code-mixing rates and language proportions of the two older siblings. Looking at the data from a similar time range (around age 8), the Sister’s average mixing on Estonian days between ages 7;7 to 8;1 was 3.7%, while the Brother’s average mixing rate between age 7;10 to 8;0 on Estonian days was 4.1%. However, the Brother used more English on Estonian days (but not vice versa). The amount of data for the Brother from this period is not sufficient to adequately evaluate whether this was a constant phenomenon or due to a lack of a reliable amount of data, but it is noteworthy that the Sister did not show any tendency to speak in English on Estonian days. It is interesting that the preferred language was the non-societal language, as it has usually been reported in other studies (see, for example, Quick et al. 2020) that the societal language is the preferred language (although in these cases there is also often more input in that language). Although English has a high status in Estonia, it is questionable to what extent a 3-year-old is able to perceive this.

In summary, although it would have been logical to assume that if the input proportions remained the same, then the language choice patterns of a young bilingual would not show much change, this was not the case. The rate of code-mixing fluctuated quite a bit during language development, and the child showed a clear preference for one language over the other for a period. However, the change in willingness to speak both languages again is notable and important for parents of bilingual children, as it suggests that parents who wish to raise their children bilingually and notice a strong language preference can try to increase the quality of input (different native speaker(s) or focused book reading or other language building activities) in that language to help the child speak it.

### 4.3 Influence of siblings on the rate of code-mixing

It has been suggested in the literature that young bilingual children adjust their code-mixing rate to their interlocutors (Comeau et al. 2003). However, previous findings have not been conclusive. To gain further insight into this matter, it was attempted to see with a statistical analysis whether the presence of older siblings, who code-mixed, affected the code-mixing rate of the younger sibling. The results did not show that the presence of older siblings affected the code-mixing rate of the younger sibling. Comparing the average code-mixing rate of two time periods where data was available showed similar

average code-mixing rates to recordings done with and without siblings. This was the case for days when the recordings were made on days the family spoke English (27% vs 29%) as well as for days they spoke Estonian (29% vs 28%). This suggests that other factors seem to play a more decisive role in a young child's code-mixing.

However, the data from siblings was sporadic as they were not the target persons for these recordings, which limited the use of different statistical measures. Perhaps a priming study, which would enable to take into account every utterance in sequence from all participants, and hence also the specific rate of code-mixing, would give a better picture of the role siblings play in code-mixing and language choice. This would also allow to investigate whether the presence of a younger sibling, who code-mixes and switches into the other language due to lacking constructions or lower entrenchment levels of given constructions, could prime the older siblings to code-mix or switch languages of the conversation.

## 5. Conclusion

This paper investigated input-output proportions and code-mixing in the speech of a bilingual English-Estonian speaker (2;3-5;0) in a situation where the input language proportions remained the same throughout the recording period. The analysis showed input effects on MLU scores as the MLU scores of both languages increased at the same rate, just as the balance of languages in the input would predict based on previous research findings. The data also revealed that when the MLU in each language reached around 2.5 the proportion of code-mixed utterances started to decrease and as the child's English and Estonian skills grew (as measured by MLU), she code-mixed less and less. However, effects of Estonian being an agglutinative language can be noted as the MLU of Estonian utterances did not increase as much (from 1.75 [age 2;3] to 2.44 [2;9] vs in English from 2.02 [2;3] to 2.82 [2;9]).

It was found that in the early stages of language acquisition the child's language proportions followed her input proportions, though she also code-mixed a lot. This is in line with previous research findings. However, at age 2;11, without any changes in the input language proportions, the child started to speak less and less Estonian, and eventually at age 3;2 only 19% of her utterances on Estonian days were in Estonian. This suggests that other factors, not solely input level, played a role in the language choice of this young bilingual. However, data from later recordings showed a turn towards the initial output balance between the languages, though more code-mixing remained on Estonian days compared to English days.

An attempt was also made to see whether the presence of bilingual older siblings, who used code-mixing in their speech influenced the younger sibling to code-mix. Although the data was limited, code-mixing by older siblings showed no influence on the younger siblings code-mixing rate. However, more data from siblings would be needed to better assess the effect, or it is suggested that a study on priming be conducted, which would allow to take into account every utterance from all participants and to analyze the influence one person's code-mixing and language choice has on the other participants in the conversation.

## References

- Ambridge, Ben, Evan Kidd, Caroline F. Rowland & Anna L. Theakston. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42 (2), 239–273.
- Baird, Piret. 2022. Enabling tool: Estonian-English code-mixing of a 2-year-old with balanced input. *Philologia Estonica Tallinnensis* 7, 80–102.

- Barron-Hauwaert, Suzanne. 2011. *Bilingual siblings: Language use in families*. Bristol: Multilingual Matters.
- Behrens, Heike. 2006. The input–output relationship in first language acquisition. *Language and Cognitive Processes* 21 (1–3), 2–24.
- Bernardini, Petra & Suzanne Schlyter. 2004. Growing syntactic structure and code-mixing in the weaker language: The Ivy Hypothesis. *Bilingualism: Language and Cognition* 7 (1), 49–69.
- Brown, Roger. 1973. *A first language: The early stages*. Harvard University Press.
- Bybee, Joan L. 2013. *Usage-based theory and exemplar representations of constructions*. (Eds.) Thomas Hoffmann & Graeme Trousdale. Vol. 1. Oxford University Press.
- Byers-Heinlein, Krista & Casey Lew-Williams. 2013. Bilingualism in the early years: What the science says. *LEARNING landscapes*. NIH Public Access 7 (1), 95.
- Cameron-Faulkner, Thea & Claire Noble. 2013. A comparison of book text and child directed speech. *First Language* 33 (3), 268–279.
- Chabon, Shelly S., Louise Kent-Udolf & Donald B. Egolf. 1982. The temporal reliability of Brown’s mean length of utterance (MLU-M) measure with post-stage V children. *Journal of Speech, Language, and Hearing Research* 25 (1), 124–128.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.
- Comeau, Liane, Fred Genesee & Lindsay Lapaquette. 2003. The modeling hypothesis and child bilingual code-mixing. *International Journal of Bilingualism* 7 (2), 113–126.
- David, Annabelle & Li Wei. 2008. Individual differences in the lexical development of French–English bilingual children. *International Journal of Bilingual Education and Bilingualism*. Routledge 11 (5), 598–618.
- De Houwer, Annick. 1990. *The acquisition of two languages from birth: A case study*. Cambridge University Press.
- De Houwer, Annick. 2009. *Bilingual first language acquisition*. Bristol: Multilingual Matters.
- Gaskins, Dorota, Ad Backus & Antje Endesfelder Quick. 2019. Slot-and-frame schemas in the language of a Polish- and English-speaking child: The impact of usage patterns on the switch placement. *Languages* 4 (1), 8.
- Gaskins, Dorota, Maria Frick, Elina Palola & Antje Endesfelder Quick. 2019. Towards a usage-based model of early code-switching: Evidence from three language pairs. *Applied Linguistics Review* 12 (2), 179–206.
- Gaskins, Dorota, Antje Endesfelder Quick, Anna Verschik & Ad Backus. 2022. Usage-based approaches to child code-switching: State of the art and ways forward. *Cognitive Development* 64, 101269.
- Genesee, Fred, Elena Nicoladis & Johanne Paradis. 1995. Language differentiation in early bilingual development. *Journal of Child Language* 22, 611–31.
- Hoff, Erika. 2006. How social contexts support and shape language development. *Developmental Review* 26 (1), 55–88.
- Hoff, Erika & Cynthia Core. 2015. What clinicians need to know about bilingual development. *Seminars in Speech and Language* 36 (2), 89–99.
- Hoff, Erika, Cynthia Core, Silvia Place, Rosario Rumiche, Melissa Señor & Marisol Parra. 2012. Dual language exposure and early bilingual development. *Journal of Child Language* 39 (1), 1–27.
- Jiménez Gaspar, Amelia & Laia Arnaus Gil. 2022. The role of (older) siblings in the acquisition of heritage languages: Early Catalan–German bilingualism in Germany. In *Cultura en transició. Estudis culturals a la catalanística*, 165–209.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Vol. 1. Stanford University Press.
- Lanza, Elizabeth. 1992. Can bilingual two-year-olds code-switch? *Journal of Child Language* 19 (3), 633–658.
- Mishina-Mori, Satomi. 2011. A longitudinal analysis of language choice in bilingual children: The role of parental input and interaction. *Journal of Pragmatics* 43 (13), 3122–3138.
- Montag, Jessica L., Michael N. Jones & Linda B. Smith. 2015. The words children hear: Picture books and the statistics for language learning. *Psychological Science* 26 (9), 1489–1496.
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge University Press.
- Nicoladis, Elena & Fred Genesee. 1997. Language development in preschool bilingual children. *Journal of Speech-Language Pathology and Audiology* 21 (4), 258–270.

- Nicoladis, Elena & Giovanni Secco. 2000. The role of a child's productive vocabulary in the language choice of a bilingual family. *First Language* 20 (58), 003–028.
- Nieminen, Lea. 2009. MLU and IPSyn measuring absolute complexity. *Eesti Rakenduslingvistika Ühingu aastaraamat* 5, 173–185.
- Noble, Claire, Giovanni Sala, Michelle Peter, Jamie Lingwood, Caroline Rowland, Fernand Gobet & Julian Pine. 2019. The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review* 28, 100290.
- Pearson, Barbara Zurer. 2007. Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics* 28 (3), 399–410.
- Petitto, Laura Ann, Marina Katerelos, Bronna G. Levy, Kristine Gauna, Karine Tétreault & Vittoria Ferraro. 2001. Bilingual signed and spoken language acquisition from birth: implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language* 28 (2), 453–496.
- Place, Silvia & Erika Hoff. 2011. Properties of dual language exposure that influence 2-year-olds' bilingual proficiency: Dual language exposure and bilingual proficiency. *Child Development* 82 (6), 1834–1849.
- Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2021. Entrenchment effects in code-mixing: Individual differences in German-English bilingual children. *Cognitive Linguistics* 32 (2), 319–348.
- Quick, Antje Endesfelder, Dorota Gaskins, Oksana Bailleul, Maria Frick & Elina Palola. 2020. A gateway to complexity: A cross-linguistic comparison of child bilingual speech. *International Journal of Bilingualism* 25 (3), 800–811.
- Quick, Antje Endesfelder, Elena Lieven, Ad Backus & Michael Tomasello. 2018. Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism* 8 (3), 393–409.
- Redlinger, Wendy E. & Tschang-Zin Park. 1980. Language mixing in young bilinguals. *Journal of Child Language* 7 (2), 337–352.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Theakston, Anna & Elena Lieven. 2017. Multiunit sequences in first language acquisition. *Topics in Cognitive Science* 9 (3), 588–603.
- Thordardottir, Elin. 2011. The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism* 15 (4), 426–445.
- Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press.
- Virsu, Minna. 2022. *Vuorojen pituudet ja kompleksisuus koodinvaihdon yhteydessä au pairin ja perheen monikielisissä keskusteluissa* [Mean length of utterance scores and utterance complexity along with code-mixing in the multilingual discussions of an au pair and a family]. M. Virsu Master's Thesis. <https://oulurepo.oulu.fi/handle/10024/20556>.
- Volterra, Virginia & Traute Taeschner. 1978. The acquisition and development of language by bilingual children. *Journal of Child Language* 5, 311–326.

Submitted April 2024  
Accepted August 2024