

RESEARCH ARTICLE

Building languages: Estonian–English two-year-old bilingual’s reliance on patterns in code-mixed utterances

Piret Baird 

School of Humanities, Tallinn University, Narva mnt 25, 10120 Tallinn, Estonia
E-mail: pbaird@tlu.ee

(Received 21 March 2023; revised 16 January 2024; accepted 18 January 2024; first published online 16 February 2024)

Abstract

This paper examines patterns in an Estonian–English bilingual child’s spontaneous speech, employing a computational application of the traceback method, which is used in usage-based linguistics. Forty-five hours of data were analyzed to check what proportion of patterns from code-mixed utterances are attested in the child’s monolingual data and in her input. Pattern overlap between the child’s and the caregivers’ speech was also examined. Results show that about one-third of code-mixed utterances can be traced back to the child’s input and one-third also to her own monolingual data. A little over half of the child’s utterances are either chunks or frame-and-slot patterns from the caregivers’ speech. These results make it evident that the traceback method can also be applied to language pairs that are genealogically more distant, though limitations exist.

Keywords: bilingualism; code-mixing; English; Estonian; language acquisition; traceback method; usage-based linguistics

1. Introduction

Recent decades have seen an increase in the application of usage-based theory in child language acquisition research. Some of the main assumptions of this theory are related to innateness and modularity. Supporters of usage-based theory reject the formalists’ claim that people are born with language categories. According to formalists and nativists, language abilities are biologically inherent and it is not possible to acquire a language with cognitive abilities (for an overview of the debate see Ambridge & Lieven 2011). However, usage-based theorists see children’s growing linguistic abilities as a collection of abilities formed from language use (Tomasello 2003, 2009). Here, it is necessary to emphasize that the input the child receives and the output they themselves produce are both important because the latter also influences the processing of the new language heard (Lieven, Salomo & Tomasello 2009). A speaker’s



continuous linguistic experience is at center stage, and one can say that the structure of language develops based on language use (Tomasello 2003).

In this paper the usage-based approach is applied to bilingual first language acquisition data, with an additional emphasis on code-mixing. Code-mixing in this paper is understood as the use of more than one language in one utterance (Cantone 2007). Code-mixing is a frequent phenomenon in language acquisition for bilingual children, and studying why children mix and how they do it has been an interest of researchers in the past few decades. With the gain in momentum of the usage-based approach, its application has also been extended to bilingual language acquisition research, with the assumption that findings from monolingual usage-based research should also be applicable to bilingual data.

This study will add an additional language pair, the genealogically unrelated languages English and Estonian, and a different input situation to confirm or rebut the earlier results (presented in Section 2.2). Bilingual children's language acquisition is not well studied in language pairs involving Estonian, hence this study will also aim to start filling that gap. The current study seeks to find an answer to three research questions: (i) What proportion of code-mixed utterances can be accounted for with the help of constructional patterns found in the child's monolingual data and in the caregiver's input? (ii) To what extent do the patterns in the child's output overlap with patterns attested in the input? (iii) Which are the most frequent frames that attract code-mixing? Answering the above questions will also enable discussion of whether the traceback method can be applied to a child's spontaneous speech data from two genealogically unrelated languages. The last question has been posed in several studies (for example Gaskins et al. 2022, Koch, Hartmann & Quick 2022), and while the current study sheds some light on that question, it was felt that answering the question properly would warrant a study of its own.

This paper is structured as follows. First, in Section 2.1, an overview of language acquisition from a usage-based perspective is given. Section 2.2 focuses on usage-based research in terms of code-mixing, and in Section 2.3 an overview of the traceback method is given. Section 3 contains information about the data and the participant. Thereafter, Section 4 covers the method of the study and Section 5 outlines the results. Section 6 contains a discussion of the results.

2. Usage-based theory and its application in language acquisition research

2.1 *Language acquisition and usage-based theory*

According to usage-based theory, children's language acquisition is based on actual usage events and takes place gradually in a piecemeal fashion. As children learn language, they build up their linguistic knowledge from the experience they gain from various usage events that involve the input they hear and the output they produce (Lieven & Tomasello 2008). Children's language mirrors their input and hence their construction development is input-dependent. Usage-based accounts reject the nativist approach's claim of the poverty of the linguistic input, according to which the input children receive is too poor for the child to infer structural principles of a language from it (Chomsky 1986). This claim has been challenged by

numerous studies (e.g. Tomasello 2003). Child-directed speech (CDS¹) has been shown to be highly repetitive, and hence not ungrammatical and chaotic as Chomsky claims, meaning that it is possible for children to extract patterns from it. Cameron-Faulkner, Lieven & Tomasello (2003) showed by analyzing the speech of twelve English-speaking mothers that the speech of mothers is highly repetitive, and there is a high correlation between the mother's speech patterns and the patterns of their two-year-old children. Also, a mother's speech includes many lexically specific patterns, such as *What do X, Are you X, It's X*. This shows that children's speech is based on actual usage events, as parents use recurring patterns and these are in turn used by the children in their early speech. This also suggests that when studying language acquisition it is also important to investigate input patterns. Such input–output correlations have been found in other languages as well, including languages that exhibit a relatively free word order (see for example Behrens 2006, Stoll, Abbot-Smith & Lieven 2009).

For learning their language(s), children use a variety of social and cognitive mechanisms such as pattern finding, intention reading, categorization, entrenchment, schema formation, generalization skills, and joint attention (Tomasello 2003, Lieven & Tomasello 2008, Behrens 2009, Schmid 2017, Schmid 2020). Humans are able to recognize patterns in language, and that is central to language acquisition (Tomasello 2003:34). When a person is acquiring a language, the mind works on categorizing and generalizing the linguistic information heard. The mind tries to recognize similarities, filter out non-recurring aspects and compare commonalities with the items that have already been stored. Each new unit is categorized based on similarities and dissimilarities with stored units (Behrens 2009). The ability to create categories by reinforcing commonalities and filtering out differences is called SCHEMATIZATION. Categorization and schematization are closely related processes of abstraction, in that they both involve comprehending that two or more entities are similar by abstracting away their differences (Langacker 2000). So children continually use the input they receive to generalize in order to learn.

Children hear constructions repeatedly in their input, and every repeated encounter with a given unit leaves memory traces. When the unit keeps recurring, the unit stabilizes and the specific linguistic unit is automated. This is called ENTRENCHMENT and it takes place along with the schematization processes (Behrens 2009). Langacker (2000) mentions degrees of entrenchment. According to him, we need to keep in mind that some units are more entrenched than others. When a schema is used frequently, its strength of representation increases, which allows it to become fully processed and entrenched in use (Schmid 2017). As a unit occurs frequently enough and does so in salient contexts, then it becomes retrievable as a whole and the speaker will not need to assemble it (Langacker 1987, 2000). The more often a given pattern occurs, the easier and faster the activation of it will be. Hence, units that are more entrenched are more likely to be used to assemble a specific target utterance (Langacker 2000, Dabrowska 2004), and therefore it could be suggested that speakers will use entrenched units in their daily speech since they are easier to process and faster to retrieve.

According to usage-based theory, children start out by learning low-scope constructions. These are form-meaning pairings of any size. Initially, children's speech is composed of mostly LEXICALLY SPECIFIC UNITS (also called FROZEN CHUNKS

or FROZEN PHRASES in the literature) that often directly imitate input. These lexically specific units can be morphemes, words, or even recurrent word and morpheme combinations. The child may not understand a given construction's internal structure (but can), and uses it as a whole with a specific meaning. These lexically specific units are frequent in early speech (Lieven, Pine & Barnes 1992).

As children's linguistic experience grows, the complexity of utterances increases and the utterances become more abstract (Dabrowska 2004, Lieven & Tomasello 2008). Children start combining a limited set of semantically related words. Such patterns are called PIVOT SCHEMAS or LOW-SCOPE FORMULAS (Braine 1963, Lieven et al. 2003, Ambridge & Lieven 2011). These have a consistent and fixed word order (*bye-bye-X*, *more-X*) and initially they co-exist with less developed strategies of word combination (e.g. *chair apple*) (Miorelli 2017). Moreover, using the above-mentioned cognitive skills, children form patterns and relationships between constructions and their parts. They build their linguistic knowledge around concrete words and phrases. Children start to analyze acquired strings and develop PARTIALLY SCHEMATIC CONSTRUCTIONS (also called FRAME-AND-SLOT PATTERNS in the literature) that have a fixed part and a slot into which novel material can be inserted. For example, when a child initially utters *I want ice cream* as a whole, he/she will dissect it into *I want X*, where *I want* is lexically fixed and *X* is a slot that can be filled productively. When this happens, schematization has occurred (Lieven et al. 2003, Lieven & Tomasello 2008).

From partially schematic units children move on in their linguistic development to FULLY SCHEMATIC CONSTRUCTIONS. Children analogize across schemas based on functional or formal similarity between them, and are then able to develop adult-like abstract constructions. This is also when children will form adult-like syntactic categories such as verb and noun (Ambridge & Lieven 2011).

Studies with monolingual children have supported the claims of usage-based theory that early child language is built up piece by piece, based on chunks and low-scope patterns. For example, Lieven et al. (2003), Dabrowska & Lieven (2005), and Lieven, Salomo & Tomasello (2009) found that they were able to identify recurrent patterns in children's speech. Moreover, it has been found that children's speech is composed of frames with open slots, into which the children add items and the complexity of the items increase in time (Lieven, Salomo & Tomasello 2009).

Hence, children's language can be placed on a continuum of schematicity from fixed chunks to fully schematic constructions, and the complexity of the language can vary. Moreover, it can be argued that children's language development is input-dependent and a gradual process (Behrens 2006). Constructions grow and change item by item, piece by piece. The development along the continuum is connected to the input the child hears, as the formation of fully and partially schematic constructions depends on their presence and frequency in the input the child hears and also on the output he/she produces. The child's speech becomes more productive as he/she learns to segment utterances, categorizes units, and generalizes them.

2.2 Usage-based code-mixing research

The usage-based approach has also gained momentum in code-mixing research. Much of the research in the past few decades has focused on finding the structural

constraints on code-mixing and suggested different code-mixing typologies (see for example Poplack 1980, Myers-Scotton 1997, MacSwan 2000, Muysken 2000, Bernardini & Schlyter 2004). These approaches ran into problems as many counterexamples for each of the proposed constraints were presented. However, usage-based approaches focus, in general, more on the cognitive mechanisms that guide code-mixing. It can be assumed that the mechanisms mentioned in the previous section that play a role in monolingual acquisition are also the driving forces behind multilingual acquisition, which has been shown to involve code-mixing (Quick & Hartmann 2021). Researchers have used data from bilingual children to see if bilingual first language acquisition is also highly formulaic and built up in a piecemeal fashion based on chunks and low-scope formulas.

Several studies have demonstrated that bilingual children's early speech is also composed of chunks and frame-and-slot patterns and is input-dependent. For example, Quick et al. (2019) analyzed the speech of a German–English bilingual (2;3-3;11) and found a large overlap between the child's code-mixed data and his input data. This suggests that the child uses a caregiver's input to form (code-mixed) utterances. Furthermore, it has been shown that there are individual differences in inventories originating from input situations. Quick & Hartmann (2021) did a cross-corpus study of two German–English bilingual children using the traceback method (explained in Section 2.3), and found considerable individual differences in the rate of successful tracebacks. This suggests that children pick up linguistic knowledge from their input, as tracing each child's data back to their own caregiver's input was more successful than tracing it back to the other child's input data. Moreover, also tracing back each child's code-mixed utterances to their own monolingual data produced more successful tracebacks compared to tracing code-mixed utterances back to the other child's monolingual data. This posits individual differences as well as the child's reliance on patterns, even when formulating code-mixed utterances.

The same constructional patterns that form the basis of monolingual speech can, according to Koch, Hartmann & Quick (2022), be assumed also to be the basis for code-mixed utterances. For example, *I want X* can be the basis for *I want **see*** (from here on, Estonian is marked in bold) 'I want it' (Fiona, 2;7) if the child has uttered the frame *I want* as well as the slot filler **see** 'it' before. Therefore, it is likely that fully lexically specific units and frame-and-slot patterns can also be found in children's code-mixed utterances. Gaskins et al. (2021) analyzed the speech of one German–English, one Polish–English, and one Finnish–English child, and they were successful in tracing back 63–65% of code-mixed utterances to monolingual data. They further analyzed the slot fillers in frame-and-slot patterns, and found that in over 90% of cases in monolingual frames the slots were filled with fillers from the other language. Their findings show that children use patterns from their input in their speech.

These frame-and-slot patterns are also the basis for productive language use in monolinguals as well as bilinguals. Partially schematic constructions are where children build their language in a piecemeal fashion. They fill the slots with novel items, and step by step increase the complexity and length of their constructions. For example, Quick, Backus & Lieven (2021) found that the German–English bilingual children in their study initially used single nouns in the slots of lexically

specific frames, but with development they added determiners and adjectives, thus gradually building their language to contain the more general and abstract category of noun phrase. Moreover, Gaskins et al. (2019) have shown in their study involving three different language pairs that bilingual children can also form their partially schematic constructions by filling the slot with material from either language, thereby sometimes producing code-mixed utterances.

One of the ways in which usage-based theory has been applied in language acquisition studies has been the traceback method, which allows identification of linguistic patterns, such as fully and partially schematic constructions, in a corpus. In the next section, an overview of the method is given.

2.3 Traceback method

The traceback method was developed by Lieven, Behrens, Spears & Tomasello (2003) and Dabrowska & Lieven (2005). The aim of the method is to show how the child's utterances are related to his/her previous utterances or to the utterances found in his/her input. The method can help establish that children learn language in an item-based way, piece by piece as claimed by the usage-based theory. The method also shows which patterns the child uses. For this purpose the longitudinal corpus documenting a child's language acquisition is divided into two parts: a main corpus and a smaller test corpus. The test corpus in the original traceback method is composed of the last one or two recording sessions. The child's utterances in the test corpus (called target utterances) are traced back to the main corpus with the purpose of finding precedents (called component units). When tracing back, two types of precedent are searched for: fixed strings and frame-and-slot patterns. First, it is checked whether a verbatim match for a given utterance from the test corpus can be found in the main corpus. If a match is found and it meets the frequency threshold, it is considered a fixed string. It is assumed that the child has that fixed string available as an entrenched unit. Most traceback studies use a frequency threshold of two occurrences (for example Dabrowska & Lieven 2005, but see Koch, Hartmann & Quick 2022 for a discussion and results of manipulating the threshold). According to Dabrowska (2014), a threshold of two is enough because the corpora only capture a very small proportion of the child's speech. Dabrowska (2014) also showed that a higher frequency threshold leads to more failed derivations, though the difference is rather small. The same finding has been confirmed by Dabrowska & Lieven (2005) and Koch, Hartmann & Quick (2022).

If no verbatim match was found, the researcher tries to search for partial matches using predefined operations. In this way the method tries to reconstruct the cut-and-paste strategy that children follow in reverse (Hartmann, Koch & Quick 2021). This is connected to the usage-based theory's assumption that children use pattern finding and categorization skills when learning a language. The number, names, and scope of the operations differ between the individual traceback studies. All traceback studies have used the operations SUBSTITUTE and ADD. The operation ADD allows a linear juxtaposition of strings (Hartmann, Koch & Quick 2021). In the first study Lieven et al. (2003) did not restrict this operation in any way, but subsequent studies set the criterion of the combination being syntactically and semantically possible in any order. This means that conjunctions cannot be used with this

operation. Furthermore, Hartmann, Koch & Quick (2021:8) note that ‘this essentially limits the application of ADD to vocatives like *mommy* or adverbials like *now* and *then*’. Koch, Hartmann & Quick (2020) explain that if the target utterance is *will das haben Mama* ‘wanna have that Mummy’, and *will das haben* can be derived and the algorithm also finds that *Mama* meets the frequency threshold, then the derivation is considered to be successful. The reason for the restriction is the desire to avoid implausible derivations (Dabrowska & Lieven 2005). In the case of the operation SUBSTITUTE, if the target utterance from test corpus is *Give me a banana*, and utterances such as *Give me a bear* and *Give me a car* can be traced back to the main corpus, then a frame-and-slot pattern *Give me a THING* can be posited (Hartmann, Koch & Quick 2021).

The use of the SUBSTITUTE operation is semantically constrained. Therefore, most traceback studies apply various semantic slot categories such as referent, process, attribute, location, direction, possessor, and utterance (Hartmann, Koch & Quick 2021). The use of semantic categories comes from the assumption that children can make semantic generalizations from the input.

Since there are occasionally cases where more than one derivation is possible, the researcher follows three principles: (i) the largest possible schema is used, (ii) the slot is filled by the longest available unit, and (iii) the minimum number of operations is used (Hartmann, Koch & Quick 2021). If during the traceback procedure no pattern was found, the derivation is considered a fail.

While the traceback method has now been applied to several monolingual datasets (for example English, Italian, German) and a few bilingual language pairs (mostly German–English, but also Polish–English, Finnish–English²), several studies (e.g. Quick & Hartmann 2021, Gaskins et al. 2022) have pointed out the need to apply the method to other language pairs, especially to those involving genealogically distant languages. Quick & Hartmann (2021) find that the method works well for German–English bilingual data because the languages are structured fairly similarly. With a language pair like Estonian and English, we are dealing with two genealogically distant languages. Estonian, a Finno-Ugric language, has a lot of inflection and freer word order than English. This could mean that the method will not be able to detect patterns or that the rate of failed derivations will be higher than in other studies. Hence, applying the method to Estonian–English bilingual data should shed some light on the applicability of the method to a genealogically distant language pair.

3. Participant and data

The participant in the study was a two-year-old simultaneous Estonian–English bilingual child who resides in Estonia. Both parents speak Estonian and English well and the family has opted to use a family language policy where they rotate the language spoken by days of the week (three days of Estonian and four days of English), thus deviating from most other bilingual language acquisition study participants, as usually the one-parent–one-language method has been employed. Most of the child’s input came from the immediate family, as she did not attend daycare before or during the data collection period. The parents recorded the child’s spontaneous speech about once a week between ages 2;3 and 2;11. Altogether 45 hours of data were recorded. There were more recordings done on days when the

family spoke Estonian (31 hours of recordings from Estonian days and 14 hours from English days). Starting from 2;5, there was at least one recording in English each month. The input also includes some data from the two older bilingual siblings (ages 7;6 and 5;3 at the beginning of recordings, who also grew up with the same family language policy) and the father, as they were present for some of the recordings. The parents did not use much³ code-mixing in their interactions with the child, but the siblings' use of code-mixing was more frequent.⁴

4. Method

I will follow Quick & Hartmann (2021) in their modification and application of the traceback method. In order to answer the research questions, three analyses were carried out. To find out the proportion of code-mixed utterances that can be accounted for with constructional patterns in the child's monolingual data, the code-mixed utterances of the child ($N = 3,265$) were the test corpus and the child's own monolingual utterances were the main corpus ($N = 4,149$, of which 1,883 in Estonian and 2,266 in English) for the first analysis. Using the child's code-mixed utterances as the test corpus allows us to see if, similarly to monolingual language acquisition data, the child's code-mixing can also be accounted for with the help of constructional patterns from the child's own output.

To assess the proportion of code-mixed utterances that can be accounted for with constructional patterns in the caregivers' input, the test corpus was again the child's code-mixed utterances ($N = 3,265$) and the caregivers' data⁵ was the main corpus ($N = 21,970$, of which 14,773 in Estonian, 6,784 in English, and 399 mixed) for the second analysis. Tracing the child's code-mixed utterances back to mostly monolingual input data from the caregivers can help us see to what extent the patterns in the child's code-mixed speech are influenced by the input she receives, which was mostly monolingual. Hence, these two analyses will show to what extent the child's code-mixed utterances are composed of fixed chunks and partially schematic utterances from her own earlier speech or that of her caregivers.

For the third analysis all of the child's utterances were the test corpus ($N = 9,342$) and all of the caregivers' utterances were the main corpus ($N = 21,970$). This analysis will seek to answer the second research question, and attempt to show to what extent the child's speech is composed of fixed chunks and partially schematic utterances that are also present in the input from caregivers.

The method was implemented computationally using the code provided by Quick & Hartmann (2021) with the modification of removing the code for the other child (which they used for the cross-corpus analysis portion of their study). As earlier traceback studies have used various operationalizations, which has made comparing the results more difficult, then using the same code allows better comparability with the study by Quick & Hartmann (2021).

The algorithm works as follows.

- For each utterance in the test corpus (which were all code-mixed for the first two analyses) the algorithm checks whether there is a verbatim match in the main corpus. If a match is found, the derivation is considered successful.

Table 1. Example of the algorithm’s search function options

__ ei taha cornflakes’e
__ do not want cornflakes
mina __ taha cornflakes’e
I __ want cornflakes
mina ei __ cornflakes’e
mina ei __
__ taha cornflakes’e
mina __ cornflakes’e

- If no match was found, it checks whether a frame-and-slot pattern can be found for the utterance. For this, up to two consecutive words are replaced by a wildcard in the search expression. This is the same as the substitute operation in the traditional traceback method. As an example, Table 1 shows the options the algorithm will check to see whether they are found at least twice in the main corpus. In Table 1 our target utterance is *mina ei taha cornflakes’e* ‘I do not want cornflakes’ (2;5).

Thereafter, the algorithm checks if the omitted words are attested in the main corpus. If this is the case, the pattern candidate is considered valid. If multiple pattern candidates are found, then the ones with the longest consecutive fixed string are preferred. For example *mina ei* __ is preferred over *mina* __ *cornflakes’e*, as in the latter pattern candidate only one word is before and after the open slot. Also, pattern candidates with utterance-initial fixed strings are preferred over candidates with an utterance-initial open slot, but longer consecutive strings are prioritized over utterance-initial patterns. If no pattern candidate fulfills these requirements, the derivation is considered to have failed.

The disadvantage of the computational implementation is the lack of accounting for semantic or syntactic information. This means that implausible patterns could be postulated. However, Quick & Hartmann (2021:4) claim: ‘there is no guarantee that the linguistically informed patterns identified in previous traceback studies are psychologically plausible’ (for a more in-depth discussion of this issue see Hartmann, Koch & Quick 2021).

5. Results

5.1 The proportion of code-mixed utterances accountable with constructional patterns

In the first analysis, I found out how many code-mixed utterances in the child’s recorded speech can be traced back to her own monolingual data. The data indicated that 31% of the child’s code-mixed utterances can be traced back to her monolingual utterances and were found to be frame-and-slot patterns (see Figure 1). In the second analysis I wanted to find out how many of the child’s code-mixed utterances can be accounted for with the help of constructional patterns found in the child’s input from caregivers. The data revealed that 31% of code-mixed utterances can

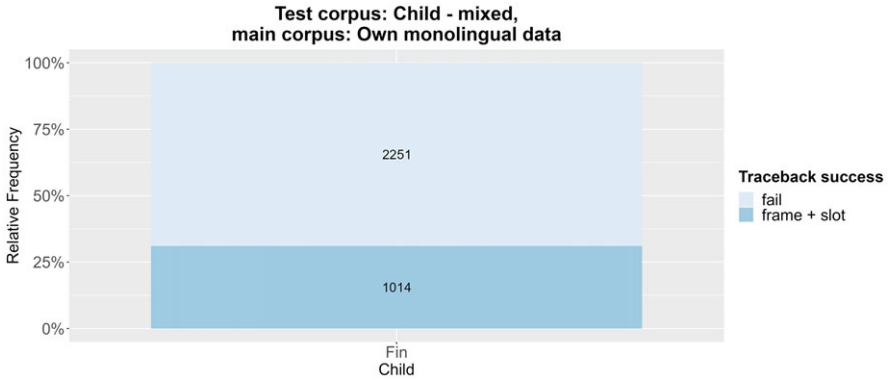


Figure 1. Traceback results: Child's code-mixed data traced back to the child's monolingual data.

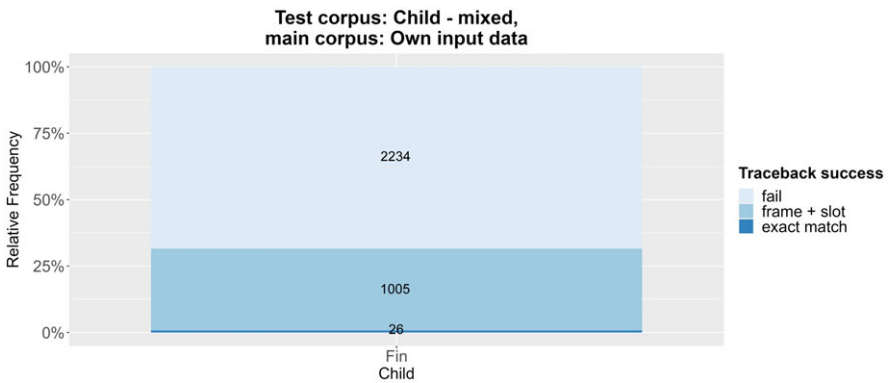


Figure 2. Traceback results: Child's code-mixed data traced back to the caregiver's data.

also be constructed from the caregivers' utterances (see Figure 2). Both of these results show that the child's code-mixed speech also seems to be constructed from patterns found in the input (her own as well as those of her caregivers).

In Figure 2, we can also see that when tracing back from the child's code-mixed data to the caregivers' data, a small proportion of utterances were verbatim matches. These were not all code-mixed verbatim matches, though some were (for example the child and her brother both say *Tell my saba* 'Tell my tail.NOM⁶'). Some of these matches were also homographs (for example the child uttered *Me ei saa-nud* 'Me no get-PST = did not get', and the input contains the same, but here, instead of the English, *me* is the Estonian third person plural *me* 'we'), but the automated system did not differentiate the languages and hence considered them a match.

5.2 The proportion of utterances accountable with constructional patterns from input

For the third analysis, when all of the child's utterances are the test corpus and the caregivers' utterances are the main corpus, then 14% of the utterances in the test

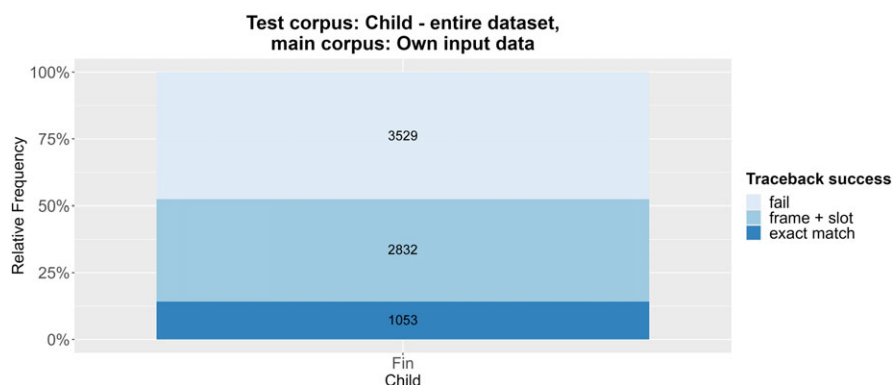


Figure 3. Traceback results: All of the child's utterances traced back to the caregiver's data.

corpus are exact matches and 38% are frame-and-slot patterns (see Figure 3). This shows that the traceback success is higher when all of the child's utterances are included in the test corpus, which was to be expected when code-mixed utterances are the test corpus and the main corpora contain no or little code-mixing, as was the case with the data used.

The results of the study confirm the findings of earlier traceback studies showing that the early Estonian–English bilingual child's speech also relies on fixed chunks and frame-and-slot patterns and the method is applicable to genealogically more distant language pairs. However, there were fewer successful tracebacks compared to most earlier studies, but the results align with Quick & Hartmann (2021), whose operationalizations I used. As outlined earlier, the operationalizations of the traceback method have varied and most of the studies have analyzed the data (semi)manually, and these factors influence the results to some extent, along with other factors related to the specific language pair under investigation. The smaller number of successful tracebacks could also be due to the size of the corpus and the freer word order of Estonian, which will be taken up further in the discussion section.

5.3 Most frequent frames

Overall, the traceback analysis detected 469 different frames. As a reminder, more data (31 vs. 14 hours) was recorded on days when the family spoke in Estonian. While the analysis detected a few more frames in Estonian (241 compared to 209 different frames in English), the distribution as to how often the found patterns were detected in the data is more aligned with the proportion of recording settings: 62% of patterns were in Estonian and 35% in English. Twenty-one of the patterns were bilingual (for example *mina* ____ *also*).

As the data comes from two somewhat different input situations, days when the family spoke Estonian and days when the family spoke English, it is important to also differentiate the data in this way. On Estonian days 64% of patterns were in Estonian and 34% in English. Most of the bilingual frames (18) also come from the data recorded on Estonian days. On English days 41% of the uttered patterns were in

Table 2. Fifteen most frequent patterns detected by the traceback method

Pattern	Frequency	Example
___ emme ‘___ mommy’	117	Look emme ‘Look mommy’
See on ___ ‘This/that is ___’	34	See on wet ‘This/that is wet’
Emme ___ ‘Mommy ___’	29	Emme heart ‘Mommy heart’
___ look	25	Issi look ‘Daddy look’
emme ___ emme ‘mommy ___ mommy’	16	Emme stop emme ‘Mommy stop mommy’
Siin on ___ ‘Here is ___’	16	Siin on baby ‘Here is baby’
___ issi ‘___ daddy’	16	Only issi ‘Only daddy’
___ siin ‘___ here’	13	One siin ‘One here’
Where is ___	11	Where is emme ‘Where is mommy’
Mina tahan ___ ‘I want ___’	10	Mina tahan candy ‘I want candy’
___ also	9	Mina also ‘I also’
Emme siin on ___ ‘Mommy here is ___’	8	Emme siin on trash ‘Mommy here is trash’
Emme ___ siin ‘Mommy ___ here’	8	Emme no snowing siin ‘Mommy no snowing here’
Emme see on ___ ‘Mommy this is ___’	7	Emme see on cheese ‘Mommy this is cheese’
___ I am done	7	Emme I am done ‘Mommy I am done’

English and 57% in Estonian. Three bilingual frames come from the data recorded on English days.

For better comparability with other similar studies, it is also important to report the child’s mean length of utterance (MLU), as it has been discussed that higher MLU leads to more traceback fails (Koch, Hartmann & Quick 2022). The MLU in words for Estonian for the entire recording period is 2.31 and for English 3.04. The MLU for code-mixed utterances is 4.16 (for a more thorough analysis of the child’s MLU covering some of the data reported in this article see Baird 2022, and for the entire period see Baird [forthcoming](#)). Here it is also important to note that the MLU in Estonian might be somewhat smaller due to the morphological differences between English and Estonian, as Estonian uses more case endings (hence shorter) while English uses pre- and postpositions (hence making utterances longer).

In Table 2 the most frequent frames are presented. Six of the most frequent frames involve the word **emme** (‘mommy’) and one **issi** (‘daddy’), indicating the importance of constructions revolving around the primary caregivers. These results also show the importance of naming things at that stage of development as the second most frequent frame is **See on** ___ (‘This/that is ___’).

6. Discussion

In this paper, I have sought to find out whether a child's code-mixed utterances entail patterns from the child's monolingual data and from caregivers' input, relying on the usage-based approach that in early language acquisition children construct their utterances, even the ones with code-mixing, from pieces of language they have heard and used before (Lieven, Salomo & Tomasello 2009). I was also interested to see what proportion of the child's utterances overlap with patterns found in the input. Additionally, I have sought to shed light on the question of whether the traceback method is suitable for analyzing bilingual child language data from two genealogically distant languages, such as Estonian and English.

6.1 *Reliance on input*

The results indicate that the Estonian–English bilingual child's code-mixed utterances can be accounted for with frame-and-slot patterns found in the child's monolingual data and in the caregivers' input. The data analysis revealed that about one-third of the code-mixed utterances can be traced back to the patterns found in the child's own monolingual speech. Also, one-third of the code-mixed utterances could be traced back to the input from caregivers. Although at first it might seem that a one-third success rate is not much, it should be noted that the method is fairly conservative because the data only covers a fraction of the child's waking time speech. This means that the method most likely underestimates the number of utterances that potentially are chunks and frame-and-slot patterns. Getting an accurate estimate of many phenomena in children's speech is challenging due to several factors. First, collecting longitudinal spontaneous speech data is difficult, as it requires the family to be consistent with recording and motivated to take time for it. As the family could decide at any given point to discontinue recording, it is also risky for the researcher as the study might not be completed. Second, transcribing children's spontaneous speech is time-consuming and hence also costly. Transcribing bilingual data is even more challenging as the transcribers need to speak both languages, and code-mixing makes transcribing even more time-demanding because the transcriber often needs to listen to code-mixed utterances many times to be sure the details of both languages have been captured correctly. Therefore, more accurate estimates of the number of chunks and frame-and-slot patterns will need to wait for further enhancements in auto-transcribing (bilingual) children's spontaneous speech.

Fewer successful tracebacks are likely also to be due to the corpus size of this study. While the number of code-mixed utterances was similar to the study of Quick & Hartmann (2021), the child's corpus in this study was respectively five and six times smaller than their corpora, and the caregivers' corpus was about eight times smaller. In the corpus of this study there were about 1,000 more monolingual child utterances than code-mixed utterances, which likely has an effect on the method's ability to find patterns. The idea that the size of the corpus is a factor influencing the traceback results has been suggested by Koch, Hartmann & Quick (2020), who summarized and analyzed the limitations of the method. Based on the data from four German-speaking children, they found that the child with the smallest test

corpus also has high traceback results while the child with the largest test corpus has low traceback results. However, upon further analyzing input and contextual factors, they concluded that corpus size is likely one of several factors influencing the number of successful tracebacks. Thus, while the smaller corpus likely influenced the results of the current study, there were likely also other potential factors in play.

Nonetheless, finding patterns for one-third of the code-mixed utterances, and finding that a little over half of the child's utterances have verbatim matches or frame-and-slot patterns from the caregivers' data, aligns with the usage-based approach, according to which children acquire their linguistic repertoire piece by piece, and that even code-mixed utterances are related to what the children themselves have said before and what their caregivers have uttered. These results further support previous findings that children's early speech is connected to their input, and this seems to be so even in a language that employs a relatively free word order. The traceback method detected that one-third of the patterns found in the child's code-mixed utterances are also in the caregivers' speech. When all of the child's utterances were the test corpus and the caregivers' utterances the main corpus, then a little over half of the utterances were traced back successfully. These findings echo the results of Cameron-Faulkner, Lieven & Tomasello (2003), who have shown that a monolingual mother's speech is highly repetitive, and children often use the same item-based phrases as their caregivers. This is even true for code-mixed utterances, which are not directly retrieved from the input as the caregiver's speech contained hardly any code-mixed utterances. Rather, during language acquisition children develop frame-and-slot patterns, and for bilingual children some of the slots can be filled with material from the other language. Code-mixed utterances tend to be creative, as the child has not heard them uttered frequently in their entirety before.⁷ This was so in the current data because most of the speech the child heard did not include any code-mixing. Although a more thorough analysis would be needed, it also seemed as if the very few instances of code-mixing in parents' speech did not prime the child, but rather *vice versa*: the child's code-mixing primed the parents to code-mix on the few occasions in the data where the parents code-mixed, as is the case in example (1).

- (1) Child: this is **lepatriinu** color
 'this is ladybug color'
 Parent: **lepatriinu** color?
 'lady bug color?'
 Parent: but looks orange to me

Hence, it could be that code-mixing is the result of creative novel utterances, which in turn takes place due to different entrenchment levels of units and constructions in the given languages. However, this would need to be studied further to analyze and elaborate these findings.

It is telling that one-third of the child's code-mixed utterances could be traced back to the caregivers' speech, as one of the drawbacks of the method is its lack of accounting for the variable word order of Estonian, and therefore it is possible that the method underestimates the number of chunks and frame-and-slot patterns the

child has. While the method was initially used mainly with monolingual English utterances, it has been employed with other languages that have a more varying word order (e.g. German, Italian). However, even though in Estonian SVO is most frequently used (Lindström 2017), the word order in Estonian is reported to be quite flexible and pragmatically sensitive⁸ (Vihman 2018). For example, while commonly a child might utter *tahan kommi* 'I want candy', the order *kommi tahan* 'candy I want' would also be correct, and placing the object in the front would serve the purpose of emphasis. Koch, Hartmann & Quick (2020) note that it is not clear how typological differences between languages affect traceback results, and the topic is complicated enough to deserve a study of its own. However, being able to successfully detect patterns in an Estonian–English bilingual child's speech does show that the method can be applied to a pair of languages that are genealogically distant. A further in-depth analysis of traceback fails would enable us to better assess the extent of the effect of freer word order causing more fails.

The interplay of languages was evident when looking a little more into word order. In several code-mixed utterances, the child would seem to be playing around with the word order. For example, she would utter *Tantti miss'ib also two* 'Tantti misses also two', and then a few utterances later she would say *emme miss'ib kaks also* 'mommy misses two also', reversing the order of the last two words: see example (2).

- (2) Child: *Tantti miss'ib also two.*
 'Tantti misses (3SG-PST) also two.'
 Child: *I miss'ib also two.*
 'I miss (3SG-PST) also two.'
 Parent: You guys are missing two.
 Parent: I am missing one.
 Parent: Now is your turn.
 Child: *Emme miss'ib kaks also.*
 'Mommy misses two also.'

In example (2) it is also very interesting that in the last utterance presented the child uses the word *kaks* 'two' in Estonian whereas she has just uttered it in English, and also the mother has, in between these two utterances, only used English, including the word *two*. This, and other similar examples found in the data, warrant the need to further study priming effects in bilingual speech.

6.2 Frame-and-slot patterns

The findings of this study further verify that frame-and-slot patterns have an important role in the code-mixing of bilingual children. The data confirmed previous research results that children use monolingual frames into which they insert material from the other language, as the analysis detected only a handful of bilingual frames. It has already been well established that monolingual children's speech contains many frame-and-slot patterns that enable the child to become more creative and productive with his/her speech (Lieven et al. 2003, Lieven, Salomo & Tomasello 2009). The frequency of frame-and-slot patterns has also been found in recent research involving bilingual children (Quick, Backus & Lieven 2018, Quick, Lieven, Carpenter &

Tomasello 2018, Gaskins et al. 2019, Quick et al. 2019). For example, Quick et al. (2018) found in their data of a two-year-old German–English bilingual child that his code-mixing is often composed of partially schematic constructions in one language and the open slot is filled with material from the other language. This was also shown by the results of the current study: many of the code-mixed utterances were frame-and-slot patterns with the frame in one language and the slot filled with material from the other language. For example, in the utterance *I want something to süüa* ‘I want something to eat’, the frame was *I want something to ____*, which was filled with the word *süüa* ‘to eat’. These frame and slot patterns are seen as a way for the child to increase utterance length and complexity. Quick and colleagues (Quick, Backus & Lieven 2018, Quick, Lieven, Carpenter & Tomasello 2018, Quick, Backus & Lieven 2021) have additionally found in their studies of German–English bilingual children that the MLU of code-mixed utterances is longer than the MLU for monolingual utterances, and code-mixed utterances are also more complex. They suggest that this is so because in code-mixed utterances the child fills the slots in partially schematic constructions with highly entrenched words from either language, depending on the degree of entrenchment of a particular unit needed in a given speech act. While this study did not focus on MLU length (but see Baird 2022 for an MLU analysis of some of the same data), it was calculated to enable comparison with other traceback studies, and the code-mixed utterances did have a higher MLU than monolingual utterances, suggesting that some of the suggestions presented above also happen with the data presented here.

Code-mixing in frame-and-slot patterns seems to provide the child with an opportunity to express himself/herself in a way that otherwise might be difficult due to lower entrenchment of lexical or grammatical constructions. Certain words or constructions in each language could, due to various factors, have lower entrenchment levels, just as Langacker (2000) stated how different words and constructions have differing degrees of entrenchment. Lower entrenchment levels mean that a given lexical item or construction is harder to activate, which in the case of a bilingual child could mean that a translation equivalent or a construction from the other language could be used instead and a code-mixed utterance would result. Certain lexical units could, in a bilingual family, often be used in only one language. For example, if the child has regular contact with an Estonian-speaking grandmother in a certain context which involves often repeating the same activities (and in this context these would be carried out in Estonian), then it is likely that some of the vocabulary and phrases revolving around it become more entrenched in Estonian than English. If the child were to later talk about something necessitating the retelling of some of those activities in English – not the language in which the recurring events take place – it might be that, due to a higher degree of entrenchment, some of the words, phrases, or even grammatical patterns would be activated in Estonian while the child speaks in English. Hence, the child would use code-mixed utterances even if the other interlocutor and the general environment would warrant monolingual speaking. This is also in line with the usage-based approach to entrenchment according to which children hear some patterns frequently, these become entrenched, and are therefore easier to retrieve. The use of code-mixing in general, and also the use of frame-and-slot patterns in the non-language of the day (e.g. English patterns on an Estonian day and *vice versa*), suggests that this is a possibility, based on the findings of this study.

In any speech act of a bilingual child, various lexical units and grammatical constructions, with different degrees of entrenchment, compete for selection. It is easier and faster to retrieve lexical units and grammatical patterns that are more highly entrenched, so it is likely that bilingual children use such selection criteria when speaking. This perhaps especially so in circumstances where on the one hand there is separation of languages (in the case of this child, according to days of the week), but on the other hand the child knows perfectly well that she is understood regardless of the language spoken and the parents do not explicitly call out code-mixing.⁹ In order to shed more light on entrenchment effects, it would be beneficial to compare the amount of code-mixing produced by bilingual children in different circumstances, where in some cases code-mixed speech is understood and accepted and in other cases only monolingualism is approved (e.g. due to participants' language abilities). Similarly, it would be helpful to compare code-mixing rates when retelling various events, which usually take place in only one of the two languages of a given bilingual (e.g. certain routines of daycare would be retold in the non-daycare language). Although such studies would be complex in terms of controlling the input differentiation of the situations, they would certainly shed further light on the interplay of languages and entrenchment effects.

Entrenchment related to input and input effects was also present in the data of the current study. The data showed that there was a similar number of different frame-and-slot patterns in both languages (241 Estonian, 209 English), which reflected the fairly evenly distributed input the child received during the recording period (and preceding it). This echoes the findings of other researchers (Slavkov 2015, Quick, Lieven, Backus & Tomasello 2018, Quick, Backus & Lieven 2021) in which the child's output language proportions followed the input language proportions. For example, Quick, Backus & Lieven (2021) noted that the output language of all three of the German–English bilingual children in their study followed their input. This was especially clear for the child whose input situation changed during the recording period, as the change was reflected in his output language proportions. The balance of frames in this study could be the result of both languages having a fairly even number of elements and constructions entrenched. In the case of one language being more dominant in the input and output, it could be proposed that in one language the child suppresses code-mixing more easily, or perhaps there is no need to code-mix because he/she has the necessary linguistic tools to produce wanted utterances monolingually.

Code-mixing seems to be constructed around frame-and-slot patterns in the following manner: the frame is activated in one language and an element from the other language fills the open slot, for example: *I do not want* + *see* 'I do not want it'. The data implies this as there were very few bilingual frames (only 21). Gaskins et al. (2019) suggest, based on their data of three language pairs, that bilingual frames contain words which are phonologically close in both languages. They point out examples from their data such as *X for mir* in the case of a German–English bilingual, as English *for* and German *für* are phonologically close, as are English *me* and German *mir*. However, the data from the Estonian–English bilingual child's schemas did not show any phonological proximity, and were similar to the examples cited in Gaskins et al. (2019) for the Finnish–English bilingual child's data whose schemas contained examples of phonological proximity as well as no phonological proximity¹⁰). For example, my data contained the following schemas: *X on also* 'X is

also', *X want see* 'X want this', *me tahtsin X* 'me wanted X'. As can be seen, none of these show any phonological proximity. This is likely due to the genealogical distance between Estonian and English, which does not facilitate code-mixing in frames as it might do in the case of genealogically close languages, suggesting also that there are other reasons for code-mixing in frames besides phonological proximity.

7. Conclusion

In this paper a computational application of the traceback method was used to examine what proportion of patterns in the child's code-mixed utterances are found in the child's monolingual utterances and in the caregivers' input data. Also, it was investigated what proportion of patterns in the child's output overlap with the patterns found in her input. The results show that about one-third of code-mixed utterances can be traced back to the child's own monolingual utterances and one-third to the input provided by the caregiver. The results also revealed that a little over half of the child's output can be traced back to verbatim matches and frame-and-slot patterns found in the caregivers' speech. The importance of frame-and-slot patterns in code-mixing became evident, as in the case of code-mixed utterances the child tended to utter a frame in one language and fill the slot with material from the other language. Hence, it can be said that frame-and-slot patterns are an attractive spot for code-mixing to take place. Also, this seems to indicate that in bilingual language acquisition different constructions are entrenched to various levels, and these constructions compete during language production. Constructions with a higher level of entrenchment are therefore likely to be used over less entrenched ones, sometimes resulting in code-mixed utterances.

The results also show that the traceback method can be applied to a language pair where one language has variable word order. The above results mean that the method can be applied to language pairs that are genealogically distant from one another and where one employs a free word order, as Estonian does. However, it should be noted that the exact effect of variable word order on traceback success remains unclear and should be studied further.

Overall, the findings of the study lend further support to usage-based theory's view of language acquisition, as the data showed input–output effects and that the child's speech was built up piece by piece.

Acknowledgements. I would like to thank the anonymous reviewers and editors of the *Nordic Journal of Linguistics* for the constructive, positive, and extremely valuable comments they provided, which helped me to improve this article. I am also grateful to my supervisor, Professor Reili Argus, and for her support.

Competing interests. The author declares none.

Notes

1 CDS = child-directed speech, MLU = mean length of utterance.

2 The study involving Polish–English and Finnish–English was carried out manually and involved a relatively small corpus (for further details see Gaskins et al. 2019).

3 The parents only used code-mixing in the recordings for certain conventionalized expressions; for example, the English-speaking grandmother was always *grandma* and the Estonian-speaking equivalent was always *vanaema*, as is said in Estonian.

- 4 About 3–6% of the siblings' speech contained code-mixed utterances; they were present for about half of the recording sessions.
- 5 The siblings' input was also included in the main corpus, as they also provide substantial input for the child, including code-mixed input.
- 6 Abbreviations used for glossing: 3SG-PST = third person singular past, NOM = nominative, PST = past.
- 7 While the bilingual siblings did use code-mixing in their speech, it should be noted that the siblings attended daycare and school and hence were not the main source of input for the child.
- 8 However, to the best of the author's knowledge, there are no studies based on spontaneous speech data which have analyzed word order in Estonian.
- 9 From Lanza's (1992) discourse strategies, the parents used mostly either minimal grasp strategy, adult repetition, or move on. However, no specific analysis was conducted as to how often each of these was used, as it would warrant a study of its own.
- 10 There were only five bilingual frames for the Finnish–English bilingual child in their data, but note here that Estonian and Finnish are both Finno-Ugric languages.

References

- Ambridge, Ben & Elena Lieven. 2011. *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.
- Baird, Piret. 2022. Enabling tool: Estonian–English code-mixing of a 2-year-old with balanced input. *Philologia Estonica Tallinnensis* 7(1). 80–102.
- Baird, Piret. Forthcoming. Language choice and code-mixing in a longitudinal study of an Estonian–English bilingual child.
- Behrens, Heike. 2006. The input–output relationship in first language acquisition. *Language and Cognitive Processes* 21(1–3). 2–24.
- Behrens, Heike. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics* 47(2). 383–411.
- 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.
- Braine, Martin D. S. 1963. The ontogeny of English phrase structure: The first phase. *Language* 39(1). 1–13.
- Cameron-Faulkner, Thea, Elena Lieven & Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6). 843–873.
- Cantone, Katja F. 2007. *Code-switching in bilingual children*. Springer.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.
- Dabrowska, Ewa. 2004. *Language, mind and brain: Some psychological and neurological constraints on theories of grammar*. Edinburgh: Edinburgh University Press.
- Dabrowska, Ewa. 2014. Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics* 25(4). 617–653.
- Dabrowska, Ewa & Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16(3). 437–474.
- Gaskins, Dorota, Oksana Bailleul, Anne Werner & Antje Endesfelder Quick. 2021. A crosslinguistic study of child code-switching within the noun phrase: A usage-based perspective. *Languages* 6(1). 29.
- 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.
- Hartmann, Stefan, Nikolas Koch & Antje Endesfelder Quick. 2021. The traceback method in child language acquisition research: Identifying patterns in early speech. *Language and Cognition* 13(2). 227–253.
- Koch, Nikolas, Stefan Hartmann & Antje Endesfelder Quick. 2020. The traceback method and the early construction: Theoretical and methodological considerations. *Corpus Linguistics and Linguistic Theory* 18(3). 477–504.
- Koch, Nikolas, Stefan Hartmann & Antje Endesfelder Quick. 2022. Traceback and chunk-based learning: Comparing usage-based computational approaches to child code-mixing. *Languages* 7(4). 271.

- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*, vol. 1, *Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 2000. A dynamic usage-based model. In Michael Barlow & Suzanne Kemer (eds.), *Usage based models of language*, 1–64. Stanford, CA: Center for the Study of Language and Information.
- Lanza, Elizabeth. 1992. Can bilingual two-year-olds code-switch? *Journal of Child Language* 19(3). 633–658.
- Lieven, Elena & Michael Tomasello. 2008. Children's first language acquisition from a usage-based perspective. In Peter Robinson & Nick C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 178–206. Routledge.
- Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30(2). 333–370.
- Lieven, Elena, Julian M. Pine & Helen Dresner Barnes. 1992. Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language* 19(2). 287–310.
- Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3). 481–507.
- Lindström, Liina. 2017. Lause infostruktuur ja sõnajärg [Information structure and word order]. In Mati Ereht & Helle Metslang (eds.), *Eesti keele süntaks* [Syntax of the Estonian language], 537–564. Tartu: Tartu Ülikooli Kirjastus.
- MacSwan, Jeff. 2000. The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition* 3(1). 37–54.
- Miorelli, Luca. 2017. *The development of morpho-syntactic competence in Italian-speaking children: A usage-based approach*. University of Northumbria at Newcastle.
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol. 1997. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Oxford University Press.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics* 18(7–8). 581–618.
- Quick, Antje Endesfelder & Stefan Hartmann. 2021. The building blocks of child bilingual code-mixing: A cross-corpus traceback approach. *Frontiers in Psychology* 12. 682838.
- Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2018. Partially schematic constructions as engines of development: Evidence from German–English bilingual acquisition. In Eline Zenner, Ad Backus & Esme Winter-Froemel (eds.), *Cognitive contact linguistics: Placing usage, meaning and mind at the core of contact-induced variation and change*, 279–304. De Gruyter Mouton.
- 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, Stefan Hartmann, Ad Backus & Elena Lieven. 2019. Entrenchment and productivity: The role of input in the code-mixing of a German–English bilingual child. *Applied Linguistics Review* 12(2). 225–247.
- 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.
- Quick, Antje Endesfelder, Elena Lieven, Malinda Carpenter & Michael Tomasello. 2018. Identifying partially schematic units in the code-mixing of an English and German speaking child. *Linguistic Approaches to Bilingualism* 8(4). 477–501.
- Schmid, Hans-Jörg. 2017. A framework for understanding linguistic entrenchment and its psychological foundations. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning*, 9–36. Berlin: De Gruyter Mouton.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Slavkov, Nikolay. 2015. Language attrition and reactivation in the context of bilingual first language acquisition. *International Journal of Bilingual Education and Bilingualism* 18(6). 715–734.
- Stoll, Sabine, Kirsten Abbot-Smith & Elena Lieven. 2009. Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science* 33(1). 75–103.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

- Tomasello, Michael. 2009. The usage-based theory of language acquisition. In *The Cambridge handbook of child language*, 69–87. Cambridge: Cambridge University Press.
- Vihman, Virve-Anneli. 2018. Language interaction in emergent grammars: Morphology and word order in bilingual children's code-switching. *Languages* 3(4). 40.