Antje Endesfelder Quick*, Ad Backus and Elena Lieven

Entrenchment effects in code-mixing: individual differences in German-English bilingual children

https://doi.org/10.1515/cog-2020-0036
Received April 9, 2020; accepted February 20, 2021;
published online March 9, 2021

Abstract: Following a usage-based approach to language acquisition, lexically specific patterns are considered to be important building blocks for language productivity and feature heavily both in child-directed speech and in the early speech of children (Arnon, Inbal & Morten H. Christiansen. 2017. The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science* 9(3). 621–636; Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press). In order to account for patterns, the traceback method has been widely applied in research on first language acquisition to test the hypothesis that children's utterances can be accounted for on the basis of a limited inventory of chunks and partially schematic units (Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3). 481–508). In the current study, we applied the method to code-mixed utterances (n = 1,506) of three German-English bilingual children between 2 and 4 years of age to investigate individual differences in each child's own inventory of patterns in relation to their input settings. It was shown that units such as *I see X* as in *I see a Kelle* 'I see a trowel' could be traced back to the child's own previous productions. More importantly, we see that each child's inventory of constructions draws heavily on multiword chunks that are strongly dependent on the children's language input situations.

Keywords: bilingual first language acquisition; code-mixing; entrenchment; individual differences; usage-based

***Corresponding author: Antje Quick**, University of Leipzig, Institute of British Studies, Beethovenstraße 15, 04107 Leipzig, Germany, E-mail: antje.quick@uni-leipzig.de
Ad Backus, Tilburg University, Tilburg, Netherlands, E-mail: A.M.Backus@uvt.nl
Elena Lieven, University of Manchester, Manchester, UK; and Division of Human Communication, Development and Hearing, Manchester, UK, E-mail: elena.lieven@manchester.ac.uk

1 Introduction

Usage-based accounts of linguistic phenomena have become more and more prevalent and perhaps nowhere more prominently so than in the study of child language acquisition. In this study, we extend this approach to acquisition in bilingual contexts. There is growing awareness that the usage-based approach has much to offer to the explanation of phenomena typical of bilingualism and language contact, including codeswitching, linguistic borrowing and change, and the simultaneous acquisition of two languages (Backus 2014; Zenner et al. 2019). The simultaneous use of two languages at the same time, usually referred to as 'codeswitching', is a prominent feature of bilingual life all over the world, and has been investigated intensively for its sociolinguistic motivations, its linguistic characteristics, and its implications for the mental representation of language. These perspectives attempt to find answers to, respectively, the questions of why people mix languages, what speech looks like when they do, and what aspects of cognition make it possible. Mixing is so ubiquitous for bilinguals that even young children who grow up in bilingual families in which each parent tends to keep the languages distinct by sticking to his/her L1 only, will spontaneously produce code-mixed utterances. This phenomenon is the subject of the current study, in which we look at how three children, each growing up in Germany with one German- and one English-speaking parent who keep the languages separate to different extents, acquire their two languages. Because of the language separation in the input, studying this kind of code-mixing in child speech will tell us something about the balance of labor between imitation and rote learning on the one hand, and creativity and productive use of learned linguistic resources on the other hand. This in turn will be useful in helping us understand how children build their inventory of linguistic units. Comparing the three children, against the background of the similarities and differences in the linguistic habits of their respective families, will help us pinpoint the differences between them, and illuminate how individual variation is manifested.

Individual variation between humans is evident in how they use their language(s). It could be seen as a paradox of communication that, on the one hand, we all have our own individual ways of talking, while on the other hand we talk similarly enough to preclude massive misunderstanding, particularly within communities in which we feel at home. Variation between speakers has been observed at many levels: pronunciation peculiarities, the words we prefer, the syntactic complexity of our utterances, our personal discourse style, etc. However, syntax has been singled out by formalist accounts of language as a domain that is claimed to exhibit very little variation: all L1 speakers are supposed to converge on

the same 'system', and it is generally accepted that it is possible to describe a general grammar that contains the (syntactic) rules of a particular language (e.g., Crain and Lillo-Martin 1999; Seidenberg 1997). We engage with this issue by presenting a case study of individual variation between the three participating children. We see it as crucial for the further development of usage-based theory that we understand how variation is manifested, since it points to the dimensions that are relatively variable in language and those that are not. In this light, syntax is an interesting linguistic domain to look at, since, for a particular language, it is not supposed to exhibit much variation in its underlying structure (e.g., Crain and Lillo-Martin 1999). The usage-based approach, on the other hand, should make us expect individual variation in the syntactic domain as well, since every language user has an individual usage profile. However, there may be additional theoretical reasons for expecting this variation to be seriously constrained. In this paper, we focus on the sub-domain of syntax we refer to as 'partially schematic constructions' (Langacker 1987); Section 2 will provide the reasons for this focus. We analyze the partially schematic constructions in the German-English mixed utterances of the children we studied, and the languages they are uttered in. We are interested in how the differences in their respective language input influence their construction of code-mixed utterances.

2 A usage-based perspective on language acquisition in bilingual settings

In the following, we consider the contribution the usage-based approach offers to the understanding of language acquisition. While the general narrative of this contribution is fairly well known (e.g., Tomasello 2003), we specifically concentrate on the importance of partially schematic constructions for building a productive grammar.

2.1 Usage-based accounts of acquisition

The usage-based approach starts out from the assumption that all human beings build up linguistic knowledge (or 'mental representations', or 'competence') on the basis of their experience. As knowledge is a property of the mind, it is structured through the learning and storing skills available to the mind. These skills are assumed to be universally shared by all human beings, so any differences are hypothesized to be caused by differences in the experiential input and perhaps

also by different abilities in using these domain-general skills (see Dąbrowska and Street 2006; Kidd et al. 2018; Street and Dąbrowska 2010). As for linguistic theory, a usage-based approach implies a bottom-up account of how people establish linguistic knowledge. As such, it contrasts with generative hypotheses which assume that syntactic knowledge is innate. In the usage-based literature, linguistic knowledge is conceptualized as an inventory of constructions (form-meaning pairings) that are organized in a network (e.g., Bybee 2010; Culicover and Jackendoff 2005; Langacker 1987). This then triggers a view of how children build up their linguistic knowledge that is very different from the generative perspective.

Usage-based work has provided evidence that children construct their grammatical systems in a piecemeal fashion by relying on general cognitive learning mechanisms such as pattern finding, imitation, and the entrenchment of frequently encountered units (Schmid 2020; Tomasello 2003). Recurrent patterns are registered and give rise to schematization, which leads to the entrenchment of schematic (or 'abstract') patterns (also referred to as 'templates', or simply as 'grammatical constructions'). Though to a lesser degree than adults, young children differ in their linguistic experiences concerning topics of conversation, the ways their interlocutors talk, and the conversational settings they get involved in (e.g., Cameron-Faulkner et al. 2003 on child-directed speech). Even if we ignore such conversational variation, every interaction is unique, and therefore all children will have different experiences with how frequently they hear particular words and structures, how particular forms have been made salient on occasions, how consistent or variable their input is, how much effort their adult caretakers put into being clear and transparent, etc. All of this implies differences in the children's linguistic experiences, and this leads to differences in, for example, vocabulary inventories, vocabulary size, grammatical proficiency, and/or discourse skills (e.g., Bates et al. 1995; Birdsong 2016; Frank et al. 2017; Hart and Risley 1995; Ibbotson 2019; Kidd et al. 2018).

While there are many studies documenting variation in lexical and communicative skills, the picture is much less clear for syntactic knowledge. Partially, this is a matter of defining the object of study. Traditionally, syntax includes only the fully abstract constructions we use to build utterances. However, one crucial aspect of usage-based linguistics is its attention to lexical-syntactic hybrids, generally referred to as 'partially schematic constructions'. These are constructions in which some of the morphemes or words are fixed (see the next section for details), and recognizing their existence complicates the relationship between lexicon and syntax. Any actual utterance may instantiate the productive use of a fully schematic construction ('syntax'), the activation of a fully memorized expression that merely instantiates that construction ('lexicon'), or the level in-between that partially schematic constructions represent. While it stands to reason

that all competent speakers of a particular language have stored the same fully schematic constructions, and therefore share the same 'syntactic knowledge', it is exactly in the inventories of these more lexically-specific constructions that a usage-based perspective would expect to find individual variation and this, in turn, might lead one to expect differences in the level of abstraction in the representations of individual speakers (Dąbrowska and Street 2006).

While it is clear that the 'final state' of language acquisition is similar enough across speakers that we can communicate with each other across a reasonably wide group of speakers of the same language, how successful this communication is will depend on the linguistic and social knowledge that individual speakers have built up over development (Schmid 2020). Frequency is suggested to be the most important factor affecting both the acquisition and the processing of constructions, with frequent structures acquired earlier and processed more readily (Bates et al. 1995; Birdsong 2016). Other things being equal (e.g., the precise nature of mappings between form and meaning), constructions with high token and high type frequency are, typically, the first learned fully schematic constructions. High token frequency leads to the entrenchment of frequent expressions, while high type frequency shows the greatest affordance for generalization (such as the transitive clause pattern in traditional grammatical description) (Ambridge and Lieven 2015; Goldberg 1995). High frequency also leads to shared inclusion in the inventories of all speakers, and thus to the relative absence of individual variation.

However, there is variation in the age at which children speaking the same language show evidence of acquiring particular, fully schematic constructions. The same holds for semantically equivalent constructions across languages. In the case of the transitive construction, input patterns both within and across languages account for these differences (see Ambridge and Lieven 2011; Chan et al. 2009; Dittmar et al. 2008). A particular example for children learning English is the evidence that they can generalize from transitive utterances with pronominal subjects and novel verbs earlier than from the same sentences if they contain full NP subjects – presumably because they hear enormous numbers of transitive sentences with pronominal subjects in their input and therefore initially develop more specific schemas for the transitive than the fully abstract construction. When we look across languages, Chan et al. (2009) have shown that English-speaking children develop the ability to understand transitive sentences with inanimate subjects (i.e., a construction that goes beyond the canonical 'animate agent-transitive verb-inanimate patient' sentences that are most frequent in their input) earlier than both German and Cantonese speaking children. The suggested explanation is that German allows fronting of the object which blocks earlier abstraction of the canonical form-meaning mapping of the SVO transitive, while Cantonese has a very high degree of subject drop, again making abstraction of the

fully schematic construction more difficult. Thus, on the way to developing a fully abstract representation of the transitive construction, children follow a route that depends on the lower-scope patterns that are prevalent in their input.

2.2 Partially schematic units

There is, by now, evidence that entrenched units can be of any size and complexity and can be schematic and/or lexically specific. This entails that there are three types of units in the inventory of linguistic knowledge that any speaker possesses (the totality of his or her 'mental representation' of language). First, there are 'lexically-specific units': morphemes, words, and recurrent word and morpheme combinations such as *What's up?*. There is a strong evidence that such multiword units are stored and processed as lexically-specific wholes (e.g., Bannard and Matthews 2008; Bybee and Scheibman 1999). The second type of unit is similar to traditional syntax: fully schematic grammatical constructions such as the English transitive (Subject-Verb-Direct Object, SVO), filled by the syntactic categories of NP VP NP, characterized as templates that have been schematized over many instantiations of specific expressions that share the same abstract structure into which words or strings of the relevant categories (e.g., NP, VP) can be inserted. Finally, there are units that have both lexically-specific and schematic parts, referred to as 'partially schematic units', e.g., *I want X*. Linguistic knowledge is therefore claimed to be more complex than just a collection of words (lexicon) and a set of abstract, fully schematic, grammatical patterns (syntax). Which of the three types of unit are instantiated in a speaker's actual utterance is, in general, impossible to say but we may assume that any adult speaker's inventory will be a combination of overlapping lexically-specific units, partially schematic constructions and fully schematic templates (e.g., Culicover and Jackendoff 2005; Diessel 2019; Ellis et al. 2016; Tomasello 2003; Wray 2002). How these are used in actual utterances will depend on factors such as lexical access, conversational priming and the degree of entrenchment of the various units in each individual's inventory.

Studies of early child language have shown that it is generally characterized as highly formulaic and that novel, 'creative' utterances are often built with a relatively small number of high-frequency frames, constructed around lexically-specific units (e.g., Lieven et al. 1997; Theakston et al. 2012). Input plays a crucial role as children may learn multi-word units directly from the input (e.g., 'want more'; Kirjavainen et al. 2009; Cameron-Faulkner et al. 2003). Evidence for this comes from some of the errors that children make. For instance, Kirjavainen et al. (2009) have shown that children's me-for-I errors can be explained by the presence

of *me* before non-finite verbs in the input (e.g., 'let me do it' in the input gives rise to 'me do' in the child's speech and 'watch me cooking' can give rise to 'me cooking'). It was also shown that this is dependent on the specific verbs that appear after 'me' in these complex sentences in the input. Lexically-specific units subsequently become analyzed and form the basis for the production of partially schematic and more complex utterances (e.g., 'want X', or 'and now I want X'; e.g., Dąbrowska et al. 2009; Diessel and Tomasello 2001). Partially schematic utterances that are used productively may give rise to more general schematization, for instance, the development of the category 'noun phrase', through the increasing schematization of determiner, adjective and noun slots (Kemp et al. 2005; Lieven et al. 2009).

Different experiences in life and different input will lead to differences in speakers' inventories. Despite the fundamental patterns of a language that linguists describe, there are different routes to productivity and each child will take its own route and building blocks (Dąbrowska 2019). By emphasizing the attainment of abstract syntax, acquisition studies in the past have often glossed over some of the individual variation in the development towards it (noted early by Bates et al. 1988; Bates and Goodman 1997). Like basic vocabulary, frequent syntactic patterns tend to overlap across all individual speakers, for reasons interesting in their own right (see below), and therefore, there will be some systematic developmental trajectories in children's acquisition of the syntactic aspects of a language. This does not necessarily hold, however, for units that are less frequently employed, such as non-basic lexicon, multi-word collocations and expressions, and also, to some extent, partially schematic constructions (Bornstein and Putnick 2012; Dąbrowska 2012). The ability to communicate with others, therefore, depends on what constructions are held in common, the speakers' ability to 'coerce' their intended meaning into the constructions available in their inventory and the hearers' ability to conduct this process in reverse, with the aid of context (e.g., Ambridge and Ambridge 2020; Langacker 2009). We would argue that this necessity to use what is available in one's inventory of constructions in order to communicate in a timely fashion, accounts for much of code-mixing in both children and adults and that differences between children's inventories will arise as a function of different input conditions.

If lexically specific units and partially schematic units are the building blocks for children's language development, it is important to have a reliable way of identifying them. This, however, presents a considerable methodological challenge. The best operationalization available so far appears to be the 'Traceback' method which has been successfully applied in various corpus studies (Lieven et al. 2009, 2003). The method was developed to test whether children's linguistic output can be attributed to a limited number of patterns which are reused with different slot fillers and thus contribute to increasing productivity and creativity.

The essence of the method is that lexically specific units and partially schematic utterances are only postulated to exist in the speaker's inventory if they can be related ('traced back') to previous utterances in the recordings. To do so, all of the utterances produced in a short test corpus are 'traced back' into the remaining data set. As such, the corpus is divided into a *test corpus* (often the last recording session) and a *target corpus* (Lieven et al. 2009). Utterances in the test corpus are "traced back" to potential precedents in the remaining corpus, the target corpus (Lieven et al. 2009: 583). The traceback approach is a strictly bottom-up approach which does not assume any *a priori* categories and only looks at the child's own output. As such, it accounts for the composition of utterances by relying on previously acquired material (e.g., Koch 2019). If utterances occur verbatim previously in the corpus, a lexically specific unit (or 'chunk') is established, e.g., *I want it*. If the target utterance is matched only partially, the procedure can yield a partially schematic unit (or 'pattern') such as [*I want* X]. The traceback method has been adopted in a number of influential studies providing support for the assumption that children's creative utterances depend on the semi-formulaic reuse of language material they have acquired and used before (e.g., Dąbrowska and Lieven 2005; Koch 2019; Lieven et al. 2009; Miorelli 2017; Vogt and Lieven 2010). Lieven et al. (2009) show that this creative filling of slots allows the child to gradually increase the complexity of form-meaning mappings in their constructions. For example, the children in this study started by using single nouns in the slots of their lexically-specific frames but this changed with development as they added determiners and adjectives, thus slowly building the more general and abstract category of 'noun phrase'. The range of slots also increased, for instance, to sub-categories of verb phrase (e.g., X-ing, X-ed). The authors suggested that the re-use of previously entrenched schemas provides the building blocks both for momentary conversational creativity (e.g., *Water bloomed these flowers, I'm going to die you*, Bowerman 1981) and for language development.

So far, the traceback method has been mainly applied to monolingual corpora (mostly English but also to German [Koch 2019] and Italian [Miorelli 2017]), but recently a number of studies have used variants of the traceback method to investigate child bilingual code-mixing (e.g., Quick et al. 2018a, 2019).

3 Bilingual acquisition and code-mixing

Bilingual children develop along the same lines and follow the same universal pathways as all other children; yet they also differ from most of the children studied in the language acquisition literature by virtue of experiencing life through two languages (e.g., De Houwer 1990; Petitto et al. 2001). While some families

strictly adhere to the OPOL (*One Parent-One Language*) regime, in other families parents use both languages and engage in code-mixing. Depending on their experiences in and out of home, some children have balanced experience with both languages, while for others, exposure and use may be skewed towards one language, to a larger or smaller extent, and this balance may shift over time. All of this potentially affects how the children acquire each language, and therefore also how their inventories of constructions develop.

There is an extensive tradition of research on children's code-mixing, much of it done within a formal linguistic perspective. The approach to investigating code-mixing has been to categorize the various types of mixing and to derive from this the universal grammatical or structural principles that might underlie mixing (e.g., Cantone 2007). Researchers working within a generative framework have, for example, tended to focus on the role of asymmetry between the languages, one being the dominant language for the speaker, and on functional elements (i.e., grammatical morphemes; e.g., Bernardini and Schlyter 2004; Petersen 1988). In mixed phrases, functional elements tend to be from the dominant language. Asymmetry between the speaker's two languages is a hallmark of studies on codes-mixing more broadly, as one common type is the insertion of content words from an embedded language into utterances that follow a grammatical frame from a matrix language (see also Myers-Scotton 1997).

Recently, some studies have applied a usage-based perspective to the analysis of code-mixing in the output of bilingual children (e.g., Quick et al. 2018a, 2018b). Again most of the cases concern children growing up in OPOL *(One parent – One language)* families, in which each parent addresses the child in their L1. This is an interesting issue because, despite the fact that children may be influenced in the degree to which they code-mix depending on whether there is a lot of code-mixing in their environment, many of their actual code-mixed utterances are likely to be 'creative', in the sense of constructed 'on the fly'. This then gives a unique insight into the relationship between the learning of entrenched patterns and their employment in the construction of novel utterances. Despite parental guidance aimed at keeping languages separate, children have been found to produce code-mixing often, and to do so mainly in partially schematic units. The predominant pattern confirmed the findings of the general literature on adult code-mixing (e.g., Myers-Scotton 2002): a word from one language fills the open slot in a partially schematic unit from the other language (Quick et al. 2018a, 2019). Quick et al. 2018a analyzed the German-English code-mixing of a German-English-Spanish trilingual child and found that 60% of all code-mixed utterances were structured around a partially schematic unit with the open slot filled by a word from the other language, e.g., *this is X* : *this is kaputt* 'this is broken'. Since monolingual utterances were shorter and less complex (even with single-word utterances excluded), the authors

suggested that code-mixing in partially schematic units functioned as a gateway to the acquisition of utterances of greater length and complexity in much the same way as suggested by Lieven et al. (2009) for monolingual children. The degree to which this is facilitated by typological similarity was addressed in Gaskins et al. (2019). In the Gaskins study, these findings regarding partially schematic units were replicated with bilingual children whose languages were less similar: English-Polish and English-Finnish. The bilingual children's code-mixed utterances also consisted to a large extent (between 61 and 63%) of partially schematic units hosting an other-language word, which contributed to the greater lengths and complexities of these mixed utterances compared to monolingual ones.

Building on the demonstrated importance of lexically specific units and partially schematic utterances in early child language, we further explore their role in the code-mixed utterances of bilingual children. Specifically, we will investigate whether or not code-mixed utterances reflect the different input conditions for the two languages. For the purpose of the present study, we used a variant of the traceback method to trace the children's reuse and development of lexically-specific and code-mixed utterances from their own previous utterances and the way in which this reflected the input situations in which they were learning their languages.

On the basis of the above, we expected the following differences to emerge from our analysis of code-mixing in the language use of the three bilingual children:

1. Since each child has a different input setting in terms of parental language distribution and language use, we expect that this will be reflected in the distribution of languages in the children's output: the more children hear a language the more often they will use it. Accordingly, children with roughly balanced input should not show these biases; This will be measured in terms of the relative number of monolingual and code-mixed utterances and of utterance lengths (MLU).
2. Concerning code-mixing, the language that enjoys the greater input should feature more often in the lexically specific parts of a code-mixed utterance, e.g., if German is more frequent in the input of a child we expect German partially schematic units to be better entrenched for that child, and code-mixed utterances to more often contain partially schematic units in which the lexically-specific part is German, e.g., *wir bauen a room* 'we build a room' based on the partially schematic unit *wir bauen X* 'we build X' (from the data of one child, Silvie, at age 3). This will be measured through the number of partially schematic units in each code-mixed utterance and the language the partially schematic units appear in, again using a variant of the traceback method.

3. If children experience a shift in their input settings this should be reflected both in the relative use of monolingual utterances and in the relative balance of their two languages in their partially schematic constructions.

4 Method

Three German-English bilingual children, Silvie, Lily and Fion, were included in the study. All three grew up with one parent who was a L1 speaker of German and the other a L1 speaker of English, all were living in a German city, and all grew up as German-English simultaneous bilinguals from birth. Despite these similarities, their input situation differed a lot.

Silvie was the only child of an English-speaking mother and a German-speaking father. Her mother was Silvie's only source of English and when the mother addressed Silvie directly she used English. Silvie also heard her mother speak German when the father was around or when they were outside their home environment and thus she knew that the mother understood and spoke German perfectly well. Additionally, the parents also converged on German as the family language since the father's proficiency in English was rather limited. As such, Silvie's only input in English came from her mother when she was alone with her or when she was addressed directly by her mother. Her German input on the other hand, came from multiple sources, such as the father, kindergarten, social environment and even her mother when she talked to the father. Silvie's mother stayed home with her child during the first, and most of the second, year and during this time Silvie was mainly exposed to English during the day. However, from 18 months onwards Silvie's input situation changed as she started attending a German kindergarten for 45 h per week and was mostly exposed to German from that time on. In short, Silvie's input situation was very German biased. The recordings were all done at home during daily routine situations, such as play time, meal situations or reading books when both parents were present. The recordings covered a span of 18 months, starting at the age of 2;4 until she was 3;9, averaging to about 2.5 h of recorded data per week (about 20 session each month) and a total of 38,220 utterances.

Our next child, Lily, was the second child of a German-English couple; her brother was two years older. Just like in Silvie's case, the mother was the English L1 speaker and the father was a German L1 speaker. However, since both parents spoke each other's language very well, they did not settle on a family language but instead used both German and English interchangeably. Lily's brother was also raised as a simultaneous bilingual and so he provided further input in German, English and occasionally in code-mixed speech. Lily stayed home during the first

year with her English-speaking mother and entered a German kindergarten at the age of 18 months for most of the day. Although Lily's input situation resembles Silvie's, her home language distribution was much more balanced. Her data covered a span of 19 months (2;3 to 3;10), and consisted of spontaneous interactions made during play situations, meals, or bedtime, mostly with both parents present. The parents recorded Lily at rather infrequent intervals, on average 1 h per week, leading to large differences in the amount of data for each month with 23,667 utterances entering the analyses.

The third child, Fion, has a mother who is a L1 speaker of German and a L1 English-speaking father. Similar to Lily, both parents spoke the L2 language quite well and they were very inconsistent in their use of their mother tongues. Fion also had an older brother and the parents and brother sometimes produced code-mixed utterances. Initially, Fion was exposed to German for most of the day, the mother stayed home with him for the first months of his life and the family lived in Germany. Fion entered a German daycare at the age of 19 months for 4 h a day and later from 24 months onwards, he went to a German-speaking kindergarten for 6–8 h per day. Until the age of three years, Fion's major input was German, therefore, but, unlike the other two participants, Fion's input situation changed after his third birthday with a shift towards more English with an extensive stay in the father's home country and more frequent visits by his English-speaking grandparents who did not speak or understood any German. The data consisted of audiotape recordings of spontaneous speech in the home environment, during playtime or meals. Most of the time both parents were present as well as his brother. The recordings covered a span from 2;3–3;11. Recordings were made on average 2 h per week, adding up to a total of 205 h for the whole period; 47,928 utterances entered the analysis.

To summarize the language backgrounds, all three children were raised in Germany, all families subscribed to the idea of an OPOL household but they differed in the degree to which they adhered to this principle (see Table 1). Silvie had only one source of English but various sources for German, and German was the family language. Lily, on the other hand, had various sources of English and German, and experienced both her parents interacting in their L2 as well. Both languages were used as the family language. Finally, Fion's English input came mostly from his father but he also frequently heard his mother and brother speaking English. Both parents were very inconsistent in their use of their L1 language. After his third birthday Fion's input to English increased thanks to frequent visits from his English-speaking grandparents and a long vacation in his father's home country.

All recordings were transcribed and coded in SONIC CHAT format (MacWhinney 2000) by a bilingual research assistant.

Table 1: Input situation for each child.

	Silvie	Lily	Fion
Mother	English	English	German
Father	German	German	English
Home language	German	German and English	inconsistent
N of child utterances	38.220	23.667	47.928

4.1 Analyses

4.1.1 Language proportions and MLU

Unlike most other studies, our recordings took place when both parents were present, resulting in a bilingual setting. Therefore, we first coded all child utterances for language type, as either English, German or code-mixed. Incomplete and unintelligible utterances were coded as "ambiguous" and excluded from the analyses. We calculated the language proportions separately per language and child and for each recording month, and we determined the MLU in words per month, for all utterances (n = 109.815), as well as separately for the English (n = 25.430), German (n = 74.345) and mixed utterances (n = 10.040). This way we could verify how the MLU of the mixed utterances compared to those of the monolingual. Most studies compute MLU's for all utterances including the mixed ones, making direct comparisons difficult. As proposed by our first hypothesis, language proportions and MLUs might reflect the input situation: the more children hear one language the higher the proportion of utterances in that language should be and the longer the utterances should also be in that language. Consequently, if there is no preference for one language over the other, proportions and MLU should not differ that much from each other.

In a second step, we analyzed the individual inventories of partially schematic units in the code-mixed utterances of these three children. For this purpose, we selected the code-mixed utterances from the first and last recording month of each child as well as the month in the middle, at age 3 (Silvie n = 631, Lily n = 554, Fion n = 321).

4.1.2 Accounting for partially schematic units in code-mixing

In order to assess the levels of schematicity involved in the children's production of utterances containing code-mixing, we used a variant of the traceback method which was designed to identify previously used entrenched words, chunks and

partially schematic utterances by relating them back ('traced back') to previous child utterances in the recordings (Lieven et al. 2003, 2009). Based on this idea, we attempted to trace back each code-mixed utterance (or parts of that utterance) to the child's own previous recordings. We therefore coded code-mixed utterances for whether they contained a lexically-specific 'chunk', a combination of chunks, a partially schematic unit or none of these. In order to do so, we first used the code-mixed utterance and traced it back to all earlier recordings. If the code-mixed utterance could be traced back at least once verbatim the utterance was coded as a chunk (see example [1] below). For example, the code-mixed utterance *und this* 'and this' was traced back to earlier recordings and it occurred at least once in the recordings of Fion and thus, it was classified as a lexically-specific, chunk. If tracing back the whole utterance was not successful, we tried to trace back parts of the utterance (examples [2] and [3]). If this was successful, utterances could potentially fall into two categories: chunk + chunk, or partially schematic units. Code-mixed utterances which contained a creative combination of more than one chunk belong to the first of these, e.g., *bist du be good* 'are you be good' uttered by Silvie, which contained two previously produced chunks *bist du* and *be good* (example 2). If just one part of the utterance could be traced back, such as *who ate all my X* as in *who ate all my schinkenwurst* 'Who ate all my ham sausage' uttered by Lily, it was coded as a partially schematic unit (example 3). Utterances which did not contain any traceable parts were coded as other, e.g., *biten sie* 'bite them' from Lily. To summarize, based on traceability of the code-mixed utterances and parts of it, four categories were distinguished:

1. Chunks: Completely lexically fixed utterances
   I möchte 'I want' (Lily 2;4)
2. Chunk + chunk: Creative combination of two or more chunks
   aber ein little one 'but a little one' (Silvie 2;4) combination of *aber ein* + *little one*
3. Partially schematic units: Lexically fixed part (chunk) with a variable open slot
   this is ein Baum 'this is a tree' (Fion 3) *this is X*
4. Other
   echt cars 'real cars'

To see whether the distribution of language in the input is reflected in what the children produce, we coded each chunk or partially schematic unit for the language it was in. If the individual language settings have an impact on the language of chunks and partially schematic units in code-mixed speech, the language with the higher proportion in the input should feature more heavily in these. For example, if one child has a predominantly German input, we expect partially schematic units to consist mostly of a fixed German part and the open slot filled

with material from the other language (*ich habe X* 'I have X'). However, if a chunk could be traced back verbatim (coding 1) it will always be a mixed chunk (*und this* 'and this'). In chunk+chunk combinations different language constellations could be encountered. For example, one chunk could be in English and the other chunk could occur in German, e.g. *I got + noch eine* as in *I got noch eine* 'I got another one'. But, a chunk in a chunk + chunk combination could also be mixed, e.g., out von + your mouth as in *out von your mouth* 'straight from your mouth'. The same distribution of languages could also occur for the lexically fixed parts in a partially schematic unit, e.g. *I want X* or *ich habe X* 'I have X' or *und this X* 'and this'.

5 Results

5.1 Input proportions and MLU

In the first step, we analyzed the language proportions and the MLUs of each language, i.e., German, English or mixed to see whether they reflected the individual input settings. Recall that each child, despite being raised as simultaneous bilinguals and living in Germany, all had different input patterns. Silvie, our first child, only heard her mother speak English but otherwise was surrounded by predominantly German input. And this is also what we see in her language proportions, throughout development she hears more German and also uses mostly German (Figure 1). Even if her mother addressed Silvie in English, Silvie mostly responded in German (German n = 28.487, English n = 5,412, Mixed n = 4,321). Lily, on the other hand, has more than one source of English input since each parent speaks the other's language very well, and the older sibling is also being raised bilingually. This input pattern also reflects her language proportions which are fairly balanced with a slight trend towards more German at the end of the recording period (Figure 1). Throughout the recording period Lily has no real preference for one language over the other (German n = 11.046, English n = 10.560, Mixed n = 2061). Fion's language proportions also reflect his input situation (Figure 1). Initially he uses more German and would be classified as a German-dominant bilingual, but recall that his input situation changed after his third birthday with a longer vacation in the father's English-speaking home country and subsequent regular visits by the English-speaking grandparents. This change is also visible in the language distribution in his data, with a shift towards more English (German n = 34.812, English n = 9,458, Mixed n = 3,658). All children hardly heard any code-mixing from their parents (<1%).

Though their largely monolingual language proportions reflect their input patterns, all children produced a sizable amount of code-mixing, usually around 8–10%.

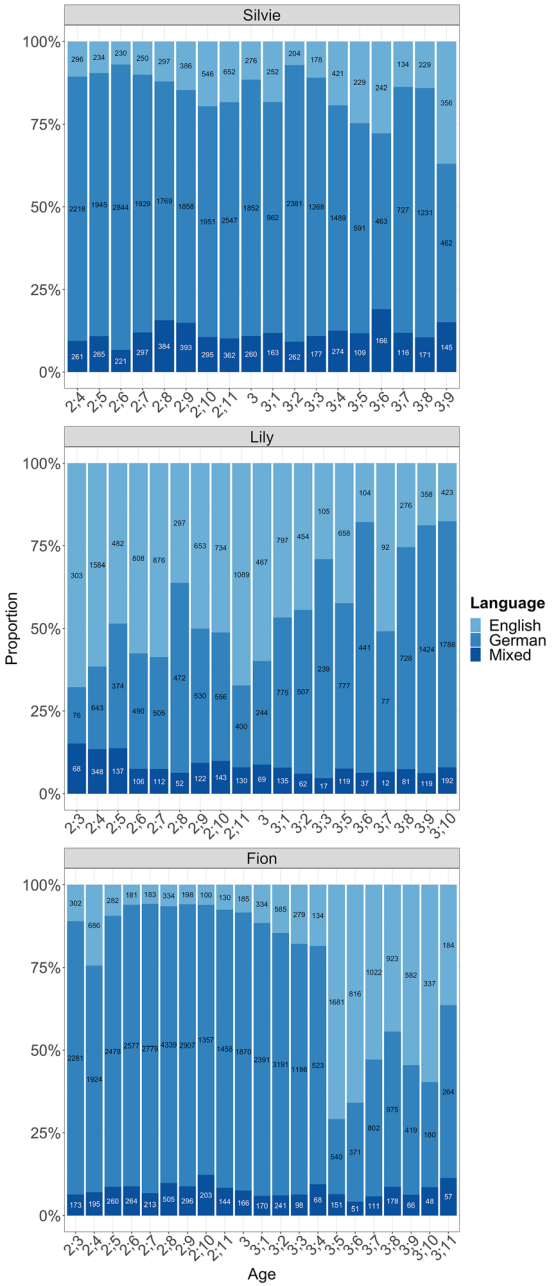

Figure 1: Language proportions for Silvie, Lily and Fion.

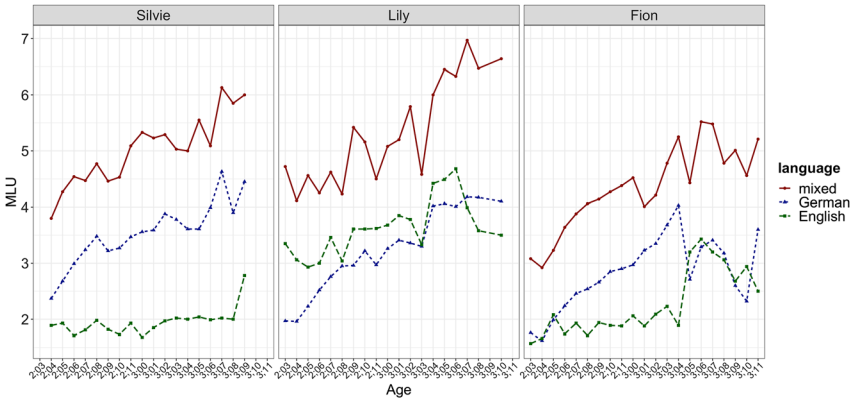

Figure 2: MLUs across age.

In the second step, we analyzed the MLUs in words for each language separately, yielding MLU scores for English, German and code-mixed utterances for each child. As expected, the monolingual MLUs largely reflect the language settings and proportions of each child (Figure 2). Silvie's German MLU gradually increases ahead of her English MLU which in turn does not become much longer for most of the recording period. This mirrors her input setting with German dominating her daily life. Lily, our more balanced child, shows monolingual MLUs which develop fairly similarly to each other. Again, this is also reflective of her input situation with parents speaking each other's language very well and thus Lily seems to profit from multiple sources of language input in the two languages. Fion's MLU develops along the lines of his input setting during his data collection period. Initially, German takes off more quickly for some time compared to English. However, his input situation changes after his third birthday with an increase in the English input and this is also mirrored by his monolingual MLUs, with English catching up at 3;5 (Figure 2).

Whereas the monolingual MLUs reflect each child's input characteristics and the language proportions in their output, an interesting picture emerged concerning the MLUs of the mixed utterances. For all three children, across the entire period of data collection, the mixed MLUs were significantly longer than the

Table 2: MLUs excluding single words.

Total MLUs excluding single words	Silvie	Lily	Fion
German	4;3	4	3;9
English	2;9	4;4	3;3
Mixed	5	5;3	4;3

monolingual MLUs (including single words in both measures), even for the stronger language ($F = 122.3$, $df_1 = 17$, $df_2 = 156$, $p < 0.001$). Even if we exclude single word utterances, mixed MLUs are still longer (although not significant anymore) than the monolingual MLUs (see Table 2). However, since all children also produced single word mixes such as *gepattet* 'patted' (Silvie at 2;4), *firewehr* 'fire ambulance' (Fion at 2;3), or *feeden* 'too feed' (Lily 2;11) we decided to include all utterances into our analyses and this raises interesting questions concerning the role code-mixing plays in syntactic development, given that an increasing MLU is generally assumed to reflect increasing syntactic knowledge.

5.2 Identifying partially schematic units in the code-mixing

In the following analysis, we investigate whether children construct their code-mixed utterances around language material they have acquired before. Therefore, our analysis is concerned with the degree to which chunks and partially schematic units are used in code-mixed utterances. In order to do that we traced back whole code-mixed as well as parts of code-mixed utterances. As explained in Section 2.1, we coded utterances as belonging to one of four categories: Chunk (completely lexically fixed) uttered exactly like this before; Combinations of two or more chunks; Partially-schematic utterances; and utterances which fit none of these categories. As can be seen in Figure 3, children significantly more often produce code-mixing in utterances built around partially-schematic constructions, such as *I want X* as in *I want die Monsterbacke* 'I want the Monsterbacke (= yoghurt brand)' ($\chi 2\,(6, N = 1{,}506) = 35.7961$,

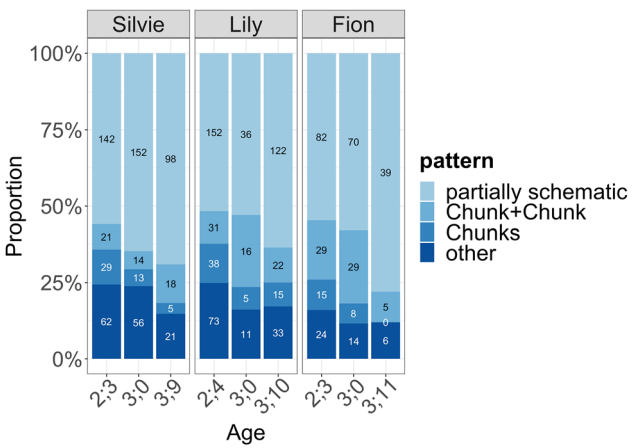

Figure 3: Schematicity in code-mixed utterances.

$p < 0.05$) than in any other category. As can be seen in Figure 3, code-mixing significantly more often takes the form of the insertion of other-language words into the open slots of partially schematic units that the traceback data suggests were already entrenched for the children. In general, most of the code-mixed utterances contained prefabricated items, chunks that had been uttered before. Note that for all three children, schematicity increases with age i.e., there is less dependence on fully lexically-specific chunks. This replicates Lieven et al.'s (2009) finding for monolingual children that 'rote' repetition decreases as language develops.

The next question concerns the language these chunks and partially schematic units were in. Chunks and partially schematic units could logically be monolingual German or English or mixed, and the question we address is whether the language of the chunks and partially schematic units reflects the input situation of the individual children (Figure 4).

As can be seen from Figure 4, the language of the chunks and the lexically specific parts of the partially schematic units was significantly more often the same as the language that dominated in each child's input and in which the MLU was higher. Given the idea that these lexically specific items with their associated open slots appear to be important building blocks for creatively combining items, even when from the other language, they seem to function as a gateway to complexity.

Silvie, our German dominant child, mostly used, in all three months analyzed, German chunks and partially schematic units to structure her code-mixing with English items filled into the open slot, e.g., *ich will jetzt X machen* 'I want to make X

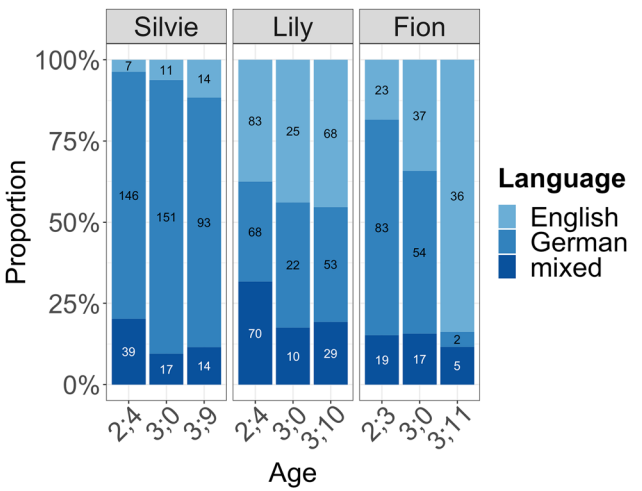

Figure 4: Language in the lexically-fixed patterns of code-mixed utterances.

Table 3: Frequent patterns.

Age/Child	Silvie	Lily	Fion
2;3/2;4	nein X (*n* = 10) 'no X'	I want X (*n* = 14)	nein X (*n* = 43) 'no X'
	da ist X (*n* = 8) 'there is'	I möchte X (*n* = 14) 'I want'	anderer X (*n* = 5) 'other X'
	ich habe X (*n* = 8) 'I have'	was is this X (*n* = 8) 'what is this'	ja X (*n* = 5) 'yes X'
	das ist X (*n* = 6) 'that/this is'	ich möchte X (*n* = 6) 'I want'	das hat X (*n* = 5) 'this has X'
	nein das ist X (*n* = 6) 'no that/this is'	das ist X (*n* = 6) 'that/this is'	X time out (*n* = 5)
3	und dann X (*n* = 10) 'and then'	I want X (*n* = 4)	this is X (*n* = 6)
	ich habe X (*n* = 13) 'I have'	but X (*n* = 3)	is this X (*n* = 6)
	ich habe schon X (*n* = 4) 'I already have'	ich bin X (*n* = 2) 'I am'	was ist X (*n* = 5) 'what is'
	hast du X (*n* = 3) 'have you'	it's X (*n* = 2)	ich will X (*n* = 5) 'I want'
	der ist X (*n* = 3) 'he is'	X in mein Zimmer (*n* = 2) 'in my room'	und this X (*n* = 2) 'and this'
3;9/3;10/3;11	und X (*n* = 6) 'and'	can you X (*n* = 20)	or X (*n* = 6)
	aber X (*n* = 6) 'but'	I want X (*n* = 9)	nee this is X (*n* = 4) no this is X
	aber ich X (*n* = 5) 'but I'	can I X (*n* = 5)	I am getting X (*n* = 3)
	und dann X (*n* = 3) 'and then'	ich möchte X (*n* = 4) 'I want'	and that's X (*n* = 3)
	kannst du X (*n* = 3) 'can you'	ich habe X (*n* = 4) 'I have'	X have to go X (*n* = 2)

now' as in *ich will jetzt the boat machen* 'I want to make the boat now', reflecting her predominantly German input ($\chi 2$ (4, N = 488) = 31,4196, $p < 0.05$). Lily, our more balanced child had no preference of one language over the other. Sometimes she used a German base, e.g., *ich bin X* as in *ich bin not* 'I am not', sometimes she structured her code-mixed utterances on an English base, e.g. *I want X* as in *I want raus* 'I want out' ($\chi 2$ (4, N = 428) = 9,4847, $p > 0.05$). Fion again, showed an interesting pattern as his shift from German to English is also projected in the lexically specific parts in his code-mixing. Whereas initially, his code-mixed utterances were based on German items (*das ein X* 'this a X' as in *das ein spider* 'this a spider'), he switched to predominantly English bases at the end of the recordings

(*now I'm getting X* as in *now I'm getting schwarz* 'now I'm getting black'), ($\chi 2$ (4, N = 276) = 63,8572, $p < 0.05$). All children also produced partially schematic utterances in which the lexically specific elements came from both languages, e.g., Fion's schema *I bin X* 'I am X' as in *I bin star wars*, or *I bin the police,* or *I bin Bob the builder*). These are interesting because there is no evidence in the input for these constructions which highlight the ways in which the children are using previously constructed code-mixes as source material for new schemas.

5.3 Individual differences in the use of partially schematic units

In our final analysis, we looked at the individual types of chunks and partially schematic units used in each child's code-mixing inventory. Since each child has its own bilingual profile, with considerable differences between them regarding their input patterns, we expect individual variation and use of partially schematic utterances at this level. In Table 3, we list the five most frequent partially schematic units for each child and month.

These lists again reflect the language proportions in the input and output, and illustrate individual differences in the use of partially schematic utterances across and within each child.

Silvie's German dominance was reflected by her frequent use of partially schematic units we may label as German constructions, by virtue of the frame being German. Lily, the balanced child, had English-based chunks and German-based chunks. Especially the first month shows an interesting pattern. She has a German-based partially schematic unit *ich möchte X* and the English equivalent *I want X* which shows that she has acquired the chunk in both languages. However, the bilingual chunk *I möchte X* is also entrenched in her repertoire which seems to indicate that she has disassembled *ich möchte* because otherwise she could not combine *I+möchte* creatively. Both chunks (*I want* and *ich möchte*) equally occur in the later months which also supports her fairly balanced proficiency in both languages, whereas the bilingual chunk *I möchte* fades out, presumably as a result of preemption (Ambridge et al. 2018), lack of reinforcement in the input (e.g., Chouinard and Clark 2003), or both.

Fion's data, again, show the shift in his input situation. During the first month of the recordings, when his input was more German, his code-mixing mainly took place in German constructions, while in his last recording, when his dominance had shifted towards English, we mainly find English schemas. At the age of three, we seem to see a transition from German towards English as English chunks start to increase in his repertoire.

Finally, we see that while some frames are used by all children, there are also many individual differences in frames they use to productively combine with words from the other language. While they use productive schemas to, by and large, do similar things with language, such as exclaiming in agreement or disaffection, expressing needs, identifying possession, and establishing the identity of things (in copula constructions), it is striking that the actual schemas that the Traceback method identifies for each of them show little overlap, at least in code-mixed utterances.

To summarize our findings, we analyzed three German-English bilingual children with differing input patterns. The differing distributions of the two languages in their input patterns were reflected in their output in a number of ways: in the proportions with which they used their languages, in the monolingual MLUs, and in the partially schematic units that framed their code-mixing. Interestingly, the code-mixed utterances had consistently longer MLUs. Individual differences were further visible in the inventory of partially schematic constructions they used productively in code-mixing, i.e., by using words from the other language to fill the open slots.

6 Discussion

In this section we summarize the results and discuss the relationship between input and pattern realization. The first main finding was that in terms of language distribution the analyses of code-mixing in the output of the three German-English bilingual children revealed that in all three cases the output reflected the individual language input situations. Silvie is in the most German-dominant environment and uses way more German in her production; Lily provides a more balanced division over the languages in accordance with her more linguistically balanced household; and Fion undergoes a shift during the period of data collection from more German- to more English-dominated output, a shift that reflects a sharp increase in exposure to English halfway during this period of time.

Second, all children code-mix, despite the fact that there is very little code-mixing in their input (<1%). Obviously, with two parents speaking different languages, they are used to a bilingual linguistic setting at home, but with parents sticking, to differing degrees, to the OPOL strategy, there is little utterance-internal mixing in the input. Yet, the children produce quite a number of such mixed utterances, as around 10% of their clauses contain code-mixing. It is important to draw attention to the fact that in OPOL settings there actually is a lot of switching between languages: both parents use their own language, and will often both be part of the interaction. This constant activation of both languages (Grosjean 2001

refers to this as the bilingual mode) is likely to make both languages active in the child's mind, and this bilingual activation promotes code-mixing.

The third main finding is that we consistently found that the MLU of mixed utterances was longer than that of monolingual utterances in either language, even compared with the language that the child was dominant in. While the MLU was consistently higher for the child's dominant language than for the non-dominant language, the fact that the average MLU of mixed utterances was even higher, suggests that mixed utterances may play an important role in driving the child's syntactic development. These findings echo those of Quick et al. (2018a), who analyzed the English-German bilingual utterances of a two-year-old child and also found that code-mixed utterances were consistently longer and more complex than the monolingual utterances. Code-mixed utterances contained more phrase and sentence level utterances and fewer fragments than the monolingual utterances. Likewise, that study also concluded that code-mixed utterances were mostly structured around partially-schematic units. These findings support the idea that (multiword) chunks and partially schematic units are integral building blocks for code-mixing and language learning in general. As such, the higher MLUs and complexities in the code-mixed utterances seem to embody another 'bilingual advantage', on the communicative level.

The final finding, we discuss relates to the properties of code-mixing. Analyzing the item-specificity in the code-mixing data, it becomes clear that at least 50% of the children's code-mixing takes place in what the Traceback method identifies as partially schematic constructions: children produce a slot-and-frame pattern in which the frame is derived from one language and the open slot is filled with elements from the other language. In these code-mixed utterances the partially schematic constructions tend to have function words as their lexically fixed parts: these form the grammatical frame, or matrix, of the construction. The language of these frames correlates with the children's input situation: the language that builds the frame is often the one that is dominant for the child. This points to some similarities with the predominant patterns of adult code-mixing found throughout the world, in which one of the languages functions as the matrix into which primarily lexical elements (words and multiword chunks) from the other language get inserted (e.g., Myers-Scotton 1997). In these approaches, code-mixing builds on the distinction between lexical and functional elements, as emphasized by both Myers-Scotton's model and other generative accounts, in the former case on grounds of psycholinguistic processing and in the latter on structural grounds. In generative accounts, including Myers-Scotton's, the explanation advanced relies on an asymmetry between the languages, one being the dominant language for the speaker. The dominant or matrix language is the language that provides the grammar while the other language may donate some content words.

From a usage-based perspective, the units of selection are not combinatorial grammatical structures and individual lexemes based on a linguistically dominant or weaker language, but words, multiword units, partial and fully schematic constructions. Unlike generative accounts as well as Myers-Scotton's model, usage-based approaches do not make a categorical distinction between lexical and grammatical aspects of language. Rather they suggest that code-mixing is determined by storage and processing characteristics that are in turn determined by social aspects of language use both in terms of learning history and momentary communicative pressures. Sometimes the cumulative result is that there is a clear base language (or in Myers-Scotton terminology a 'matrix' language), since, in this case, all partially schematic constructions will derive from the same, psycholinguistically dominant, language. However, if we consider the units of selection from a cognitive point of view we can shift away from a syntactic focus and instead explain the realization of partially schematic units based on differing levels of activation, i.e., entrenchment (Schmid 2020). Whereas Myers-Scotton draws on the assumption of a matrix language and the different roles that functional and lexical elements play in code-mixed utterances, usage-based accounts predict that children use stored multiword sequences to construct code-mixed utterances based on frequency and consequently their levels of entrenchment (see Bannard and Matthews [2008] on the retrievability of multiword units). The theoretical contrast is between a dominant language in which all the grammatical 'work' is done by this dominant language and the use of previously entrenched schemas, usually, but not always, from one language. From this point of view, the 'matrix language effect' in code-mixing, can therefore be seen as an epiphenomenon arising from the entrenchment of partially schematic constructions.

We suggest that the underlying factors that promote code-mixing are bilingual lexical selection and the use of those grammatical patterns that are most easily activated, i.e., the ones that have the higher incidence of use and are therefore likely to be better entrenched. Tomasello (2000: 77) refers to this productive use of partially schematic constructions as a "cut-and-paste" process which provides the basis for learning and for productively using language (see Dąbrowska 2014; Wray 2002). Partially schematic constructions are not just to be found in child language acquisition but they are also integral building blocks in adult language use (see Culicover and Jackendoff 2005). Patterns make execution faster and less effortful and their levels of entrenchment influence the ease of activation (e.g., Arnon and Snider 2010). In their production and comprehension of language, adults show sensitivity to distributional properties of chunks and multiword units, too. Deeply entrenched, or routinized patterns, are activated, uttered and executed without close monitoring (e.g. Havron and Arnon 2021).

In order to account for such patterns in early child language, the Traceback method has been developed. It is a conservative method, identifying units only

where the data allow it. Experimental studies, such as Bannard and Matthews (2008), do confirm that children construct partially schematic templates to comprehend and construct their utterances but it is clear that more research is needed to see how much they rely on them in relation to fully specific multiword chunks and fully schematic templates. The main advantage of the Traceback method, in our view, is its considerable face validity: it does not rely on pre-defined linguistic categories, nor on data not produced by the very child being studied and his or her caregivers. Therefore, chunks are only identified on the basis of the earlier occurrence of the exact same sequence in the output of the same child, and no assumptions of underlying structural abstractions are made. From corpus work we know that children's early language is structured around a small number of high-frequency chunks and partially schematic utterances such as *Its a X* (Lieven et al. 1997). And this is also what we found in the code-mixing of all three German-English bilingual children. In code-mixed utterances, children activated a partially schematic unit in one language and filled the open slot in another language. Our findings resonate with studies of the degree to which children show uptake from child directed speech. Parents often use utterance-initial multiword chunks which in turn are also often used in their children's early productions (Cameron-Faulkner et al. 2003). Recently, a study has shown that a German-English bilingual child also extracted patterns from the input and used these patterns in his code-mixing (Quick et al. 2018b). For example, prior the code-mixed utterance *I see a Kelle* 'I see a trowel' the father said 'I see a helicopter' providing the partially schematic unit *I see a X* which the child took up in his code-mixed utterance.

Taken together, our findings provide us with a window on language planning and language processing in young children that does not emerge as clearly from monolingual data. Code-mixing in partially schematic constructions appears to provide the child with opportunities to express content that would be difficult to say in completely monolingual utterances, perhaps because the necessary lexical competence is not in place yet.

Many of the partially schematic constructions that host code-mixing in our data consist of German functional elements that co-occur with an English noun, or vice versa. The language predominantly framing these constructions tends to be the one that is dominant for the particular child: most of Silvie's constructions have German fixed elements; Lily has constructions in both languages, and in Fion's data we see a shift from predominantly German to predominantly English constructions. This means Silvie activates the occasional English word while speaking German but much less so the other way around. Lily shows this kind of activation of words from the other language no matter whether she is currently speaking German or English. Fion activates many English words while speaking German at the start of data collection, but towards the end of when data collection mostly

activates German words while speaking English. Thus, what we are dealing with here is on-line activation on the basis of previous learning and entrenchment. There is plenty of evidence to suggest that young children of this age know perfectly well which language they are speaking and can adjust to the languages of their interlocutors (e.g., Genesee et al. 1996). The extent to which they code-mix may well be influenced by the amount they hear both languages, the extent to which both are a feature of their household environment and the knowledge they have of whether whoever they are talking to understands both languages. And, depending on how their environments change, they may well gradually cease to code-mix as each language becomes more entrenched and available or continue to do so as a function of who they are interacting with and their relative competence with each language. As they grow older, lexically-specific and partially schematic frames will remain in the linguistic representations of the individual but will also become more or less available for reanalysis as a function of different kinds of experience with language (e.g., with literacy and increasingly complex cognition, see Dąbrowska and Street 2006)

For children in the early stages of language development, tracing the origins of their linguistic knowledge and the basis on which they produce novel utterances (as code-mixed utterances often are) is more feasible. This has produced the influential body of usage-based literature on language acquisition (e.g., Tomasello 2003) that our study can be considered part of. Its specific contribution is its elucidation of the role of partially schematic units in driving syntactic development, here made visible in the creative insertion of words from the other language. As such, we have shown that the usage-based reach can be extended successfully to the investigation of code-mixed utterances.

References

Ambridge, Ben, Libby Barak, Elizabeth Wonnacott, Colin Bannard & Giovanni Sala. 2018. Effects of Both preemption and entrenchment in the retreat from verb overgeneralization errors: Four reanalyses, an extended replication, and a meta-analytic synthesis. *Collabra: Psychology* 4(1). https://doi.org/10.1525/collabra.133.

Ambridge, Ben & Chloe Ambridge. 2020. The retreat from transitive-causative overgeneralization errors. In Caroline F. Rowland, Anna L. Theakston, Ben Ambridge & Katherine E. Twomey (eds.), *Current perspectives on child language acquisition*, 113–130. John Benjamins Publishing Company. https://doi.org/10.1075/tilar.27.05amb.

Ambridge, Ben & Elena Lieven. 2011. *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Ambridge, Ben & Elena Lieven. 2015. A constructivist account of child language acquisition. In Brian MacWhinney & William O'Grady (eds.), *Handbook of language emergence*, 478–510. Hoboken, NJ: Wiley Blackwell.

Arnon, Inbal & Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62(1). 67–82.
Arnon, Inbal & Morten H. Christiansen. 2017. The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science* 9(3). 621–636.
Backus, Ad. 2014. Towards a usage-based account of language change: Implications of contact linguistics for linguistic theory. In Robert Nicolaï (ed.), *Questioning language contact. Limits of contact, contact at its limits*, 91–118. Leiden/Boston: Brill.
Bannard, Colin & Danielle Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19(3). 241–248.
Bates, Elizabeth, Inge Bretherton, Lynn Snyder, Marjorie Beeghly, Cecilia Shore, Sandra McNew, Vicki Carlson, Carol Williamson & Andrew Garrison, et al. 1988. *From first words to grammar: Individual differences and dissociable mechanisms*. New York, NY, US: Cambridge University Press.
Bates, Elizabeth, Philip Dale & Donna Thal. 1995. Individual differences and their implications for theories of language development. In Paul Fletcher & Brian MacWhinney (eds.), *Handbook of child language*. Oxford: Basil Blackwell.
Bates, Elizabeth & Judith Goodman. 1997. On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes* 12(5–6). 507–84.
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.
Birdsong, David. 2016. Dominance in bilingualism: Foundation of measurement, with insights from the study of handedness. In *Language dominance in bilingualism. Issues of measurement and operationalization*. Cambridge: Cambridge University Press.
Bornstein, Marc H. & Diane L. Putnick. 2012. Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology* 48(2). 477–491.
Bowerman, Melissa. 1981. The child's expression of meaning: Expanding relationships among lexicon, syntax, and morphology. *Annals of the New York Academy of Sciences* 379: 172–189.
Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics* 37. 575–596.
Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge University Press.
Cameron-Faulkner, Thea, Elena Lieven & Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6). 843–873.
Chan, Angel, Elena Lieven & Michael Tomasello. 2009. Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English, *Cognitive Linguistics* 20(2). 267–300.
Chouinard, Michelle M & Eve V. Clark. 2003. Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30(3). 637–670.
Crain, Stephen & Diane Lillo-Martin. 1999. *An introduction to linguistic theory and language acquisition*. Oxford: Blackwell Publishers.
Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler syntax*. Oxford University Press on Demand.
Dąbrowska, Ewa. 2019. Individual differences in grammatical knowledge. In *Cognitive linguistics – Key topics*. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110626438-012.

Dąbrowska, Ewa. 2014. Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics* 25. 617–653.
Dąbrowska, Ewa. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2(3). 219–253. https://www.jbe-platform.com/content/journals/10.1075/lab.2.3.01dab.
Dąbrowska, Ewa, Caroline Rowland & Anna Theakston. 2009. The acquisition of questions with long-distance dependencies. *Cognitive Linguistics* 20(3). 571–597.
Dąbrowska, Ewa & James Street. 2006. Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences* 28. 604–615.
Dąbrowska, Ewa & Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16. 437–474.
De Houwer, A. 1990. *The acquisition of two languages from birth: A case study*. Cambridge: Cambridge University Press.
Diessel, Holger. 2019. *The grammar network. How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
Dittmar, Miriam, Kirsten Abbot-Smith, Elena Lieven & Michael Tomasello. 2008. German children's comprehension of word order and case marking in causative sentences. *Child Development* 79(4). 1152–1167.
Ellis, Nick C., Ute Römer & Matthew Brook O'Donnel. 2016. *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Chichester, UK: John Wiley.
Frank, Michael, Mika Braginsky, Daniel Yurovsky & Virginia Marchman. 2017. Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language* 44(3), 677–694.
Gaskins, Dorota, Maria Frick, Elina Palola & Antje E. Quick. 2019. A usage-driven analysis of naturalistic code-switching data from three bilingual 2-3-year olds with English as one of their two languages and Polish, German, or Finnish as the other. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2019-0030.
Genesee, Fred, Isabelle Boivin & Elena Nicoladis. 1996. Talking with strangers: a study of bilingual children's communicative competence. *Applied Psycholinguistics* 17. 427–442.
Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
Grosjean, François. 2001. The bilingual's language modes. In Janet L. Nicol (ed.), *Explaining linguistics. One mind, two languages: Bilingual language processing*. Blackwell Publishing.
Hart, Betty & Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore, MD, US: Paul H Brookes Publishing.
Havron, Naomi & Inbal Arnon. 2021. Starting big: The effect of unit size on language learning in children and adults. *Journal of Child Language*. https://doi.org/10.1017/S0305000920000264.
Kemp, Nenagh, Elena Lieven & Michael Tomasello. 2005. Young children's knowledge of the determiner and adjective categories. *Journal of Speech, Language and Hearing Research* 48(3). 592–609.
Kidd, Evan, Seamus Donnelly, H Morten & Christiansen. 2018. Individual differences in language acquisition and processing. *Trends in Cognitive Sciences* 22(2). https://doi.org/10.1016/j.tics.2017.11.006.
Kirjavainen, Minna, Anna Theakston & Elena Lieven. 2009. Can input explain children's me-for-I errors? *Journal of Child Language* 36. 1091–1114.

Koch, Nikolas. 2019. *Schemata im Erstspracherwerb. Eine Traceback-Studie für das Deutsche*. De Gruyter.
Langacker, Ronald. 2009. Cognitive (Construction) Grammar. *Cognitive Linguistics* 20(1). 167–176.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, Vol. I. Stanford: Stanford University Press.
Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3). 481–508.
Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30. 333–370.
Lieven, Elena, Julian Pine & Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24(1). 187–219.
MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
Miorelli, Luca. 2017. *The development of morpho-syntactic competence in Italian-speaking children: A usage-based approach*. Newcastle upon Tyne: Northumbria University. http://nrl.northumbria.ac.uk/id/eprint/39990.
Myers-Scotton, Carol. 1997. *Duelling languages: Grammatical structure in code-switching*. Oxford: Oxford University Press.
Petitto, Laura Ann, Marina Katerelos, Bronna Levy, Kristine Gauna, Karine Tétrault & Vittoria Ferraro. 2001. Bilingual signed and spoken language acquisition from birth: Implications for mechanisms underlying bilingual language acquisition. *Journal of Child Language* 28(2). 1–44.
Quick, Antje Endesfelder, Elena Lieven, Malinda Carpenter & Michael Tomasello. 2018a. Identifying partially schematic units in the code-mixing of an English and German speaking child. *Linguistic Approaches to Bilingualism* 8(4). 477–501.
Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2018b. 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*, 279–304. Berlin, Boston: De Gruyter.
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* 0(0). https://doi.org/10.1515/applirev-2019-0027.
Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system. Usage, conventionalization and entrenchment*. Oxford: OUP.
Seidenberg, Mark S. 1997. Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275(5306). 1599–1603.
Street, James & Ewa Dąbrowska. 2010. More individual differences in Language Attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120. 2080–2094.
Tomasello, Michael. 2000. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 74. 209–253.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
Vogt, Paul & Elena Lieven. 2010. Verifying theories of language acquisition using computer models of language evolution. *Adaptive Behavior* 18(1). 21–35.
Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge University Press.

Eline Zenner, Ad Backus & Esme Winter-Froemel (eds.). 2019. *Cognitive contact linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change*. (Cognitive linguistics research, Vol. 62). Berlin: De Gruyter Mouton.