Running head: PARTIALLY SCHEMATIC UNITS IN CHILD CODE-MIXING

Identifying partially schematic units in the code-mixing of an English and German speaking child

Antje Endesfelder Quick[1], Elena Lieven[2], Malinda Carpenter[3,4] & Michael Tomasello[4,5]

[1]University of Leipzig

[2]University of Manchester

[3]University of St Andrews

[4]Max Planck Institute for Evolutionary Anthropology

[5]Duke University

Acknowledgements

Many thanks, first and foremost to 'Tim', the subject of this study. We are also very grateful for the hard work put in by the transcribers, Tim Manning, Henriette Zeidler, and Christiane Schwier.

Elena Lieven is co-Director of the ESRC International Centre for Language and Communicative Development (LuCiD) at The University of Manchester (www.lucid.ac.uk). The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

Please send correspondence to: Elena Lieven, The ESRC International Centre for Language and Communicative Development (LuCiD), Coupland 1 Building, University of Manchester. M139PL. E-mail: elena.lieven@manchester.ac.uk

Abstract

Intra-sentential code-mixing presents a number of puzzles for theories of bilingualism. In this paper, we examine the code-mixed English-German utterances of a young English-German-Spanish trilingual child between 1;10 – 3;1, using both an extensive diary kept by the mother and audio recordings. We address the interplay between lexical and syntactic aspects of language use outlined in the usage-based approach (e.g. Tomasello, 2003). The data suggest that partially schematic constructions play an important role in the code-mixing of this child. In addition, we find, first, that the code-mixing was not mainly the result of lexical gaps. Second, there was more mixing of German function words than content words. Third, code-mixed utterances often consisted of the use of a partially schematic construction with the open slot filled by material from the other language. These results raise a number of important issues for all theoretical approaches to code mixing, which we discuss.

Keywords: code-mixing, English – German, bilingual child, usage-based, partially schematic constructions

Identifying partially schematic units in the code-mixing of an English and German speaking child

Research into bilingualism has always been concerned with the nature of the representations underlying the ability to speak two languages with native fluency. It is clear that, among bilingual adults, the extent to which languages are kept separate in terms of code-mixing either across discourse turns or within utterances is determined by a wide range of mainly social factors, such as what the prevailing context allows, who they are talking to and whether one of their languages is dominant (e.g. Grosjean, 1998). Although the underlying linguistic representation that bilinguals have in both languages largely matches that of monolinguals in each language, it is well-documented that mixing does occur in both adult and child bilinguals.

It seems that, from early on, the two languages of young bilingual children are relatively 'encapsulated' (Paradis, 2000; Genesee, Nicoladis, & Paradis, 1995). Thus, children are proficient at adjusting to the language of their interlocutor (Genesee, Boivin, & Nicoladis, 1996), can show equivalent development to monolinguals in each of their languages (De Houwer, 1990; Meisel, 1990; Meisel, 1994; Petitto, Katerlos, Levy, Gauna, Tétrault, & Ferraro, 2001) and generally show relatively low levels of intra-sentential code-mixing (Deuchar & Quay, 2000; Müller et al., 2015; but see Comeau, Genesee, & Lapaquette, 2003).

Lexical mixing is usually explained in terms either of 'gap filling' where the speaker does not have the relevant word in the language s/he is speaking or 'activation' where for some reason the content word is more highly activated from the other language (perhaps because of frequency of use). Claims of functional mixing are more interesting because they suggest that two grammars are somehow in operation at the

same time and much research has been devoted to explicating this (e.g. Gawlitzek-Maiwald & Tracy, 1996).

Code-Mixing in Bilingual Adults and Children

Interest in code-mixing was already well developed in the 1970s and 80s (e.g. Joshi, 1985; Pfaff, 1979; Poplack, 1980; Woolford, 1983). Initially starting from a sociolinguistic perspective, researchers became interested in the underlying structures and rules governing code-mixing. Most of the research investigating code-mixed utterances has been concerned with the issue of structural constraints; all with the aim of finding universal switching points from one language to the other and possible explanations in the basic syntactic architecture of language (e.g. Cantone, 2007; Di Sciullo, Muysken & Singh, 1986; Eichler, 2011; MacSwan, 1999; Myers-Scotton, 1997; Poplack, 1980; Paradis, Nicoladis & Genesee, 2000).

A particularly influential model is the *Matrix Language Frame model* (Myers-Scotton, 1997; Myers-Scotton & Jake, 2001) which assumes that code-mixing occurs in a "frame" or "matrix" with elements of the other language embedded. Thus, the MLF is based on the asymmetry of the two participating languages and, resulting from this hierarchical ordering, the different realization of functional elements (Joshi, 1985; Myers-Scotton, 1997; Myers-Scotton & Jake, 2001). For example, Myers-Scotton (1997) makes a distinction, for each discourse situation, between the 'matrix' and the 'embedded' language, thus assuming an asymmetry between the two participating languages in a mixed utterance. One language is providing the frame and this is the language that determines word and morpheme order and that all functional elements are drawn from. This is formulated in two principles: the *Morpheme Order Principle* states

that word order must adhere to the matrix language, and the *System Morpheme Principle* says that system morphemes, a 'syntactically relevant' subclass of functional morphemes, must derive from the matrix language. Note that, like the approaches outlined above, this research also makes a principled distinction between different categories relevant to the grammar: lexical and functional (system) morphemes. The activation of system morphemes from one language is hypothesized to be an indicator of the underlying frame or 'syntactic glue' (Myers-Scotton & Jake, 2001) from which the sentence 'frame' is drawn.

Concerning child bilingual acquisition, often the approach to code-mixing is concerned with constraints that meet the requirements of the two developing grammars. Very often code-mixing is analyzed in relation to unequal language proficiencies and, resulting from that, code-mixing is often found to be directional (e.g. Bernardini & Schlyter 2004; Cantone, 2007; Eichler, 2011; Petersen, 1988). In this approach, if children are mixing, it is because of the need to fill lexical and grammatical gaps in their weaker language. For example, in an early study, Petersen (1988) suggested the *Dominant Language Hypothesis* and defined the dominant language as the one containing fewer mixed utterances. She predicted that grammatical morphemes of the dominant language could occur with lexical morphemes of either language but grammatical morphemes of the weaker language could only occur with lexical morphemes of the weaker language. A related approach is to measure the child's proficiency in each of the languages and to predict that the language in which the child is more advanced should show less mixing and, where mixed utterances do occur, the syntax should come from the stronger language.

In line with this argument Gawlitzek-Maiwald and Tracy (1996) suggested the *Bilingual Bootstrapping Hypothesis,* in which bilingual children are thought to bootstrap from the stronger into the weaker language to fill syntactical and lexical gaps. The stronger language therefore supports the weaker language. This notion is similar to the *Ivy Hypothesis* (Bernardini & Schlyter, 2004), in which the stronger language (determined by MLU and Upper bound[1]) is thought to provide the functional skeleton of the CP onto which the weaker language grows like ivy. However, there are considerable problems with all these approaches, as illustrated by the large body of counterexamples for each one of the constraints proposed. Firstly, there is evidence that children do in fact mix functional elements from the weaker language (e.g. Müller et al., 2015). Second, all these approaches assume a clear divide between functional and lexical elements and the different roles they may be playing in relation to hypothesized language dominance or the matrix language of the sentence (e.g. Petersen, 1988: no mixing of functional elements of the weaker language into the stronger language; Myers-Scotton & Jake, 2001: late outsider system morphemes must come from the matrix language). However, these approaches depend on clear definitions of dominance or of the matrix language and otherwise can be in danger of becoming circular. Finally, the great diversity across bilingual individuals is difficult to include into anyone's set of these constraints. There is no one set of constraints that can account for the range of diversity observed across bilingual children. However, variation is important as most of the time bilingual children show considerable variability not only in their development but also in their production of code-mixed utterances and this variability needs to be explained by every theory.

Individual variation is a crucial feature of the language learner partly because input and, consequently, frequency is different for every child (e.g. Fenson et al., 2007). This leads to individual bilingual profiles and proficiencies for each child. Children differ in how they learn language and also how fast they learn their language(s). These differences depend on a multitude of factors such as amount and type of input, diversity of syntactic structure, SES and individual processing abilities (Fenson et al., 2007; Hart & Risley, 1995; Hoff, 2003; Hoff-Ginsberg, 1998; Huttenlocher, Vasilyeva, Cymerman & Levine, 2002; Marchman & Fernald, 2008). Bilingual development is further influenced by age of acquisition (simultaneous or successive bilinguals), native language input, number of interlocutors, social attitudes, language typology and the amount of parental code-mixing (Comeau et al., 2003; De Houwer, 2007; Goodz, 1989; Hoff & Place, 2012; Pearson, Fernández, Lewedeg & Oller, 1997). It remains unclear how these factors relate to individual differences in code-mixing.

Therefore, in this study we wanted to depart from classical structural approaches to code-mixing and instead analyzed the code-mixed utterances based on chunking and entrenchment processes suggested by a usage-based approach.

Functional approach to code-mixing

Usage-based theory rejects the assumption of a syntax lexicon divide; language use is the combination of units of form which are associated with a specific meaning (e.g. Bybee, 2010; Tomasello, 2003). Units of form can vary in their level of complexity, ranging from individual words to large units of processing and consequently can be characterized as a taxonomy ranging from completely lexically fixed (e.g. *How are you?*), to partially schematic (e.g. *I want x*), to wholly schematic

(e.g. *NP VP NP*). Partially schematic constructions serve as a link between the two poles of lexically-fixed (item-based) and fully schematic constructions in that they consist of a fixed part and a schematic (open) part. Essentially this feature of units blurs the distinction between lexicon and syntax, functional and grammatical elements (Culicover & Jackendorff, 2005; Demiçay & Backus, 2014; Langacker, 1987). Within code-mixed utterances this means that the units of selection are not grammatical patterns and individual lexemes, but words, fixed multiword units and partially schematic constructions. Code-mixing results when a partially schematic construction is selected that is from a different language than what precedes it, or when one or more open slots in that construction are filled with lexemes or morphemes from the other language. Consequently, instead of analyzing which language provides the base and which language is mixed in, code-mixing can be analyzed in terms of the output, thus not assuming that one language is providing the general frame. What on the surface can be seen and analyzed as code-mixing is often the use of a partially schematic construction from one language and the open slot filled with material from the other language.

A further feature of an analysis in terms of form-meaning units is that language and thus grammar can be seen as a dynamic system of emergent interconnected structures that are constantly changing with usage (e.g. Bybee, 2010). This means that units can move along this taxonomy through repeated usage and become lexically more fixed. This allows for the entrenchment of schemas, including some in which the fixed part consists of words from both languages.

Sampling

One difficulty with studies of code-mixing is that code-mixed utterances are often rather infrequent. Code-mixing has been shown to have an occurrence of less than 5% on average (e.g. Müller et al., 2015). What level of sampling is required depends on what is being studied (Tomasello & Stahl, 2004) but with relatively infrequent phenomena such as code-mixed utterances, this presents major methodological challenges both in collecting a large enough sample and for subsequent analysis. Of course, sampling is a major issue in all research on children's language development whether monolingual or bilingual. Most studies of monolingual acquisition use a sampling regime of 1-2 hours of recording at best every 2-3 weeks, which, on a rough calculation, probably results in sampling about 1% of what the child says (Lieven, Behrens, Speares, & Tomasello, 2003). Child language researchers have employed a number of different methods to try to get around this problem. One is to use much denser sampling regimes, for instance of 5-10 hours per week. This can provide an estimated 10-20% of what the child says and hears but obviously is extremely costly in terms of research time and with very few exceptions can only be conducted over relatively short periods of time (Lieven & Behrens, 2011). A second is to conduct experiments and while there has been some attempt to do this for code-mixing (Comeau et al., 2003; Quick, Lieven & Tomasello, 2016), the methodology for this is in its very early stages. A final approach is that of diary collection. The disadvantage of diary studies is that, almost by definition, they are confined to a very small number of children (often only one), they are limited by the ability of the caregiver to write down what the child says and, relatedly, to cover enough of the child's waking hours to get good coverage. There have, of course, been a number of diary studies of bilingual

language development of which those by Vihman (1985) and Deuchar (1999) are notable. Vihman (1985) for example analyzed the mixing behavior of an Estonian-English bilingual child between the age of 1;1 and 2;10 and found that the child mixed primarily English function words with Estonian content words which she attributed to phonological and morphological factors. Diary studies can be extremely useful under certain circumstances, particularly if they are focused on a tightly defined phenomenon and can also get very good coverage. Examples in monolingual development are those by Bowerman (1988) and Rowland (2007). Bowerman attempted to write down every argument structure generalization that she heard her two daughters make (e.g. Water bloomed these flowers; I'm going to die you) – and, in doing so provided data that has had a major influence on the field (e.g. Tomasello, 2003). However her methodology does mean that we do not know how frequent these over-generalization errors were relative to the rest of her daughters' utterances and we know that she could not get anything like full coverage because she was not with them all day and every day. A more recent diary study by Rowland built on Bowerman's methodology and attempted to deal with these problems. Rowland collected nearly every wh-question produced by her daughter, Lauren (L), over a period of 8 months, both herself when she was with L and by persuading all adult members of her family who were caring for L to write these down. She also made recordings of L during this period. This meant (a) that she had coverage of an estimated 80% of L's wh- questions from the diary and, at the same time, had recordings from which she could conduct other quantitative analyses (for instance of the proportion of wh-questions and of the use of the same verb in wh-questions and other constructions). The present study uses this methodology: a diary

study of a large proportion of the code-mixed utterances of one child together with a relatively dense regime of recordings.

In this paper we investigate the types of code-mixing this child made to determine (a) how they compare to monolingual utterances produced by this child in the same context and (b) the extent to which they can be analyzed in terms of the multiword units and partially schematic constructions proposed in usage-based theory.

Method

Participant

The data were collected from a boy, 'Tim', between the age of 1;10 to 3;1. The family lives in Germany. Both parents are academic researchers with PhDs and Tim is their only child. The mother speaks only English to Tim. The father is bilingual in Spanish and Catalan with fluent English as an L2, and, at the time, spoke mostly Spanish to Tim with some English. The parents speak English to one another at home. From 0;5, Tim had a German babysitter for 10 hours per week and at the age of 1;1, he started attending day care in a German kindergarten. From 1;3, he spent 30-38 hours in the kindergarten per week. Both parents have a very basic knowledge of German, and the mother has a very basic knowledge of Spanish. Since Tim is also exposed to Spanish he is technically growing up trilingual. However, he had a very reduced input of Spanish during the period of this study (the mother stayed home alone with him all day until he went to kindergarten), the parents speak English with each other, and although the father often spoke Spanish to Tim, Tim usually responded in English. This is also reflected in the very few mixed utterances which contained a Spanish word: 9%

of the mixed diary utterances contained some Spanish (29 utterances) and only 3 (less than 1%) Spanish mixed utterances were found in the English recordings. Thus, for the purposes of this study we concentrate exclusively on Tim's use of English and English-German code-mixing.

Data

A diary was kept by the mother from 1;10. She is a research developmental psychologist, with long experience of both observational and experimental methodologies. However, her research does not involve either syntactic or bilingual development and she is not acquainted with the theoretical debates in the field.

The aim of the diary was simply to catch 'all' of Tim's mixed utterances heard by the mother beyond what could be recorded. All diary utterances occurred with the mother or the parents in an English context. The mother wrote down all the new mixed utterances that she heard the child say, omitting ones that were frequently repeated. The last entry analyzed in the present study is at the age of 3;1. The diary consists of 295 code-mixed utterances: 277 German-English mixes and 18 monolingual English utterances either with German word order or some other feature of German. Twenty-two of these were collected before 2;6, 168 in the months of 2;6 and 2;7, and 87 between 2;8 to 3;1. The mother reports that she made no attempt to either encourage or discourage mixes and usually did not correct them at this age.

Audio recordings were also made about halfway through the diary collection between the ages of 2;6 – 2;7. The recordings were made at home over a period of six weeks, giving a total of 30 hours of recordings. Recordings were thus made in an English context only. Recordings usually took place during playtime, meals and getting ready for bed or kindergarten. All recordings were transcribed and coded in SONIC

CHAT format (MacWhinney, 2000) by an English – German bilingual research assistant who had a basic knowledge of Spanish.

Analysis

Basic measurements

All multiword utterances in the recordings were coded for language type and utterances containing Spanish were excluded. There were thus four categories: English monolingual, German monolingual, mixed and ambiguous. Several measures were taken:

1. Proportions of mixed utterances were calculated for the recordings.
2. MLU and the mean length of the 5 longest utterances (UB5) were calculated in words in the recordings separately for the English and German monolingual utterances as well as for the mixed utterances. Ambiguous utterances and unintelligible utterances were excluded.
3. MLU and UB5 for the mixed diary utterances were calculated.

Construction types. All utterances in both the diary and the recordings (separately for the monolingual and code-mixed utterances) were coded into construction types to investigate whether the constructions for each language differed in complexity (see Table 1). A second bilingual research assistant coded 10% of the diary and 10% of the data from the recordings and reliability was high (diary, Kappa: .89; recordings: Kappa: .85).

Insert Table 1 about here

Identifying partially schematic units in code-mixed utterances

All mixed utterances in the recordings (*n*=1331) and the diary (*n*=277) were analyzed in the following way:

a) Proportions of utterances containing content words, function words or both from the non-context language,
b) Level of schematicity on the basis of the individual output of the child by grouping the code-mixed utterances into chunks (completely lexically fixed) and partially schematic constructions (utterances containing a lexically fixed part and an open slot) or belonging to neither category for the recordings which, of course, contained monolingual as well as code-mixed utterances.

Identification of fixed chunks and open slots in partially schematic constructions was based on the 'traceback' method which is used to identify how closely children's novel utterances can be related to their previous utterances (e.g. Lieven et al., 2003; Lieven, Salomo & Tomasello, 2009). Further, this analysis is also aimed at investigating whether lexically fixed chunks exist alongside partially (abstract) schemas with an open slot. Thus, defining chunks and schemas was supported by more than one occurrence of that specific string in the previous output of the child. Essentially these recurring strings can form the fixed part in a partially schematic construction. Thus, constructions were coded into the following categories:

a) Completely lexically fixed (chunks), e.g. ***hilf*** *me*[2] 'help me', which have no open slot at all
b) Creative utterances which are composed of more than one chunk, e.g. *let's* ***kaputt machen*** 'let's break it' where both chunks (*let's* and ***kaputt machen)*** are recurring units in the child's output.
c) Partially schematic constructions with an open slot, e.g. ***ich*** *want apple* 'I want an apple' was categorized as ***ich*** *want x* schema.
d) Utterances with no schemas or chunks were classified as other, e.g. ***ein*** *open* ***Mama*** 'one open Mama'.

Multilingual chunks were supported by exact repetitions by the child within the code-mixed data, e.g. ***ich*** *want* 'I want', ***und*** *this* 'and this', ***hilf*** *me* 'help me'. Monolingual chunks forming part of the code-mixed utterances were identified via previous occurrences in the monolingual data. For example, ***Mama auch*** *kick it* 'Mama kick it too' contains two chunks, ***Mama auch*** and *kick it*. Each chunk was identified by previous occurrences in the monolingual data, e.g. ***Mama auch*** occurred as exact repetition ***Mama auch***, as well as in other combinations such as ***und Mama auch ein*** 'and Mama one too', ***und Mama auch*** 'and Mama too'. *Kick it* was also identified as a chunk since it occurred in *I kick it* and *kick it.*

Partially schematic units were identified via other occurrences (at least one previous occurrence) of that schema with a variable slot. For example, the occurrence of ***ich*** *want x* 'I want x' was identified by additional occurrences with a variable slot x, ***ich*** *want this* 'I want this', ***ich*** *want* ***meine Nucki*** 'I want my pacifier', ***ich*** *want two* 'I want two'. If a schema was monolingual we also looked for exact occurrences in the monolingual data.

For example ***ich kann nicht*** *x* 'I cannot x' was supported by occurrences in the monolingual utterances, ***ich kann nicht*** as well as by occurrences in the code-mixed data, e.g. ***ich kann nicht das*** *wings* 'I cannot the wings', ***ich kann nicht*** *climb up* 'I cannot climb up'.

Finally, utterances which did not fit into either of these coding schemes or which occurred only once were classified as other.

Results

We first report quantitative analyses on the recordings before moving on to a more qualitative analysis of the mixed utterances from the recordings and the diary.

Quantitative results

There were 5783 utterances in the 30 hours of recordings. The largest proportions of Tim's utterances were context-appropriate, consisting of English monolingual utterances (66%, *n*=3785). There were 8% (*n*=483) of German monolingual utterances. Notable was his high use of mixed utterances: 23% (*n*=1331) of the data consisted of code-mixed utterances, which is far more than usually reported (e.g. Müller et al., 2015). There were only 3% (*n*=184) of ambiguous utterances.

Mean Length of Utterance

Table 2 shows the MLU and UB5 for the monolingual and mixed utterances in the recordings as well as the MLU and UB5 for the corresponding months of 2;6 and 2;7 in the diary. The overall MLU for the recordings was 3.7. MLU was also calculated for monolingual German and English utterances as well as mixed utterances. Results

showed that English monolingual MLU was 3.6 which is higher than the German MLU of 2.8. However, the mixed MLU of 4.1 (4.4 in the diary) exceeded the monolingual utterances. Thus, mixed utterances had the highest MLU while German monolingual utterances (which did not match the context) had the shortest MLU.

Insert Table 2 about here

Constructions in the recordings and diary context

Table 3 presents the relative proportions of constructions in all of Tim's utterances in the recordings, separately for the monolingual and code-mixed utterances as well as for the 277 code-mixed, diary utterances. Interestingly, the greater complexity of code-mixed utterances (identified via MLU) is also reflected in the construction analysis. A large amount of the code-mixed data in both the recordings and diary consisted of sentence level utterances (37% recordings, 50% diary). In contrast, the monolingual utterances in the recordings contained fewer sentence level utterances (26% English monolingual, 25% German monolingual).

Insert Table 3 about here

Mixing in the recordings and diary

The main pattern of mixing for the recordings and the diary was of German function words, either alone, for example, ***die*** *big kids hurt me* 'the big kids hurt me'

(76% recordings, 61% diary) or with German content words ***ich*** *want meine Nucki* 'I want my pacifier' (13% recordings, 19% diary). Mixes only of content words, such as *again* ***Strassenbahn*** 'again tram' were 11% in the recordings and 20% in the diary. In both the recordings and the diary, the most frequently mixed German function words were pronouns and determiners. A considerable proportion of all mixed utterances in the recordings (26%) contained the pronoun ***ich*** 'I' as the only mixed element e.g. ***ich*** *want water* 'I want water'. If the VP contained an English verb, the subject position overwhelmingly was filled with ***ich*** in 63% of the cases and *I* in only 7%. Seventeen percent of the mixed utterances contained only a German determiner. The German indefinite articles ***ein***, ***eine*** 'a'-MASC.SING or NEU.SING, 'a'-FEM.SING were the most frequent ones e.g. *This is* ***ein*** *lady bug* 'This is a lady bug'. Thirteen percent of the mixed utterances contained the German conjunction ***und*** 'and'.

In our next analysis we were interested in recurring units in the child's code-mixing. We therefore coded the data into fixed and open parts using the schematic constructional analysis outlined in the Method section.

Partially schematic units in code-mixing –Recordings

A major part of our analysis was concerned with the identification of chunks and schemas in code-mixed utterances. Analyses of the code-mixed utterances showed that for 88% (*n*=1178) we were able to identify chunks and schemas with at least one previous occurrence (73% of the chunks and schemas had at least two previous occurrences). Only 11% (*n*=153) of the data had no previous chunk or schema and was thus classified as other (Figure 1). Our first analysis was concerned with utterances which consisted only of a chunk. Eighteen percent (*n*=239) of the mixed utterances

were identified as complete chunks such as, ***und*** *this* 'and this' (*n*=77), ***und*** *this* ***auch*** 'and this too' (*n*=22), *this is* ***meins*** 'this is mine' (*n*=8) or ***hilf*** *me* 'help me' (*n*=6). Second, we identified utterances which consisted of two chunks strung together. Eleven percent (*n*=143) of the mixed utterances fell into that category such as ***ich*** *go* *in there* 'I go in there', ***ich*** *look* *next page* 'I look at the next page' or ***ich*** *want* *this one* 'I want this one'. Our third analysis was concerned with the identification of schemas. Sixty percent (*n*=796) of the mixed utterances were classified as partially schematic. These utterances consisted of a schema and an open slot, e.g. ***ich*** *want water* 'I want water' ***ich*** *want* sleep 'I want sleep' or ***ich*** *want my* sock 'I want my sock' which were coded as the schema ***ich*** *want x* 'I want x'.

Insert Figure 1 here

Having identified schemas, our next step was the analysis of the language of these schemas (*n*=1178) as German, English or mixed (Figure 2). Analyses revealed that more than half of the code-mixed data (56%, *n*=655) consisted of mixed schemas, for example ***ich*** *want x* 'I want x' or ***und*** *this is x* 'and this is x'. Twenty-seven percent (*n*=323) of the schemas were German monolingual, e.g. ***ich kann nicht*** *x* 'I cannot x' and 17% (*n*=200) were English monolingual, e.g. *it's x*.

Insert Figure 2 here

Summarizing the analyses of schemas showed that 88% of the data contained at least one chunk (i.e. a string that had occurred previously in Tim's speech), and, interestingly, 56% of these were mixed, such as ***ich*** *want x* 'I want x'.

In our next analysis we were interested in the language of the open slot in the mixed utterances. Figure 3 shows that if English was the language of the schema, the open slot was filled with either mixed or German material, such as *it's* ***eine*** *baby bottle* 'it's a baby bottle' or *I'm* ***alle*** 'I'm done'. However, utterances with a German schema had overwhelmingly an English open slot, for example, ***Mama auch*** *laugh* 'Mama laugh too', ***Mama auch*** *juggle* 'Mama juggle too'. Mixed chunks occurring alone (e.g. ***und*** *this*) accounted for 18% (*n*=239) of the data. In utterances containing a mixed schema, the slot was usually filled with English material, e.g. ***ich*** *look snowman* ' I look at the snowman', ***ich*** *want my sock* 'I want my sock'.

Insert Figure 3 here

Apart from identifying whether code-mixed utterances contained schemas we were also interested in the types of schemas Tim used. Analyses of types of schemas showed that Tim used a restricted set of recurring structures (Figure 4). As noted above,

his most frequent schemas were mixed in nature as in: ***ich*** *want (x)* 'I want (x)' (*n*=143**),** ***und*** *this (x)* 'and this (x)' (*n*=105), ***und*** *this is (x)* 'and this is (x) (*n*=28), ***und*** *this* ***auch*** *(x)* 'and this too' (*n*=27), ***ich*** *x it* 'I x it' (*n*=21), ***ich*** *go (x)* 'I go (x) (*n*=25). Tim's general mixing pattern was of German pronoun ***ich*** 'I' together with an English verb (*n*=236 recurring combinations). German monolingual schemas occurred more often than English monolingual schemas and were of the following type: ***ich auch*** *x* 'I too x' (*n*=21), ***ich kann nicht*** *x* 'I cannot x' (*n*=13), ***und du*** *x* 'and you x' (*n*=12), ***und das*** *x* 'and that x' (*n*=10). We further identified English monolingual schemas such as *this is x* (*n*= 26), *no x* (*n*=46), *I want x* (*n*=9), *this one x* (*n*=5).

Insert Figure 4 here

In summary, a large number of the code-mixed utterances could be classified as partially schematic constructions. However, interestingly, most of the time the schematic part contained material from both languages (***ich*** *want (x)* 'I want x). Other than this, the most notable thing about Tim's mixed utterances is that he produced translation equivalents for almost every construction.

Discussion

In this study we applied a usage-based analysis to child bilingual code-mixing and analyzed the data based on chunking processes and multi-word combinations from a 2-year-old English-German bilingual child.

First, quantitative measures showed that code-mixed utterances had the highest MLU, which was also supported by construction type analyses in that code-mixed utterances were more complex than monolingual German/English utterances. Second, the rate of mixing in the recordings was relatively frequent; we discuss below two reasons why this might be the case. Finally, a high proportion of these code-mixes could be traced back to either fixed chunks or partial schemas. We discuss the implications of these results in terms of the factors that may be involved in this child's unique combination of partially schematic constructions and novel utterances.

Why is this child different compared to previous studies?

The amount of code-mixing in this child exceeds what has usually been reported (e.g. Müller et al., 2015). This raises the question of why this child seems to be mixing so much more than other children. Two possible explanations, which are not mutually exclusive, are sampling considerations and parental strategies.

First, a strength of this study is the highly dense corpus. Whereas most studies record, for example, for an hour every fortnight (e.g. Cantone & Müller, 2005), data in this study were collected on average 5 days a week for at least one hour. Combined with the diary study, this has the advantage of capturing infrequent phenomena such as code-mixing and lowering the risk of underestimating productivity. Naturally, infrequent phenomena require a denser level of sampling to be able to have a reliable estimate of their frequency and enough data for analysis (Tomasello & Stahl, 2004). So it is possible that the dense sampling allowed us to capture a more realistic estimate of this child's code-mixing than would have been the case with, for instance, recordings for one hour per week. The dense recordings also allowed us to have sufficient utterances to

allow the possibility of tracing antecedent utterances in the corpus and identifying those that were partially schematic.

Second, parental strategies in response to code-mixing may well have influenced Tim's code-mixing. The parents in this study did not correct Tim's code-mixing and continued with the conversation, which is in line with Lanza's (2004) *Move on* strategy. Thus, code-mixing seems to have become a communicative strategy for Tim. At home Tim had only two people with whom he regularly spoke English and they responded to and accepted anything that they could understand. In contrast, in the kindergarten he heard many different people speaking German, not just two, and the teachers and children did not speak English so they would have been far less accepting of mixes which would not have been understood. This may also have contributed to entrenching a number of very frequent and common German constructions used in the kindergarten context.

Pattern of code-mixing

Analysis showed that the child mainly mixed German function words. Mixing function words has also been reported by other researchers (e.g. Lanza, 1992; Redlinger & Park, 1980; Vihman, 1985). Vihman (1985), for example, examined functional mixings in terms of phonological and morphological complexity and showed that the Estonian-English bilingual child in her study showed a preference for mixing English function words in comparison to Estonian function words. Some Estonian function words were more difficult to pronounce and morphologically more complex than English function words, and Vihman suggested that this could explain the pattern of mixing in that the child had a preference for English function words. In the present

study, however, Tim shows the reverse pattern. Although German function words are morphologically more complex - determiners, for example, need to be marked for gender and case - Tim showed a preference for these as well as for ***ich*** over *I*. However, German function words also have a higher phonological weight - compare ***ich*** and *I*, ***der*** and *the* - which might suggest that they are more salient due to typological and phonological differences. This might give them a higher degree of activation than their translation equivalents, meaning that they might be accessed more quickly.

Other researchers have explained functional mixing in relation to imbalanced languages (e.g. Bernardini & Schlyter, 2004; Cantone, 2007; Eichler, 2011; Petersen, 1988). In this approach, children mix because of unequal proficiency, and functional elements play a different role to lexical elements in these mixed utterances, since they reflect the syntactic organization of the utterance. For example, Petersen (1988) suggested that functional elements of the weaker language cannot co-occur with lexical elements from the stronger language, they can only occur with lexical morphemes of the weaker language.

However, since we cannot compare proficiency in a German context, we cannot draw a conclusion concerning the relative balance of Tim's languages. Here we explore an alternative perspective for analyzing code-mixed utterances by taking constructions as the basic components of language production.

Partially schematic units- A usage-based approach to code-mixing

Analyses of the code-mixed utterances showed that schemas in the code-mixing data of the child were relatively consistent. These schemas were very often multiword combinations which contain both functional and lexical elements (e.g. ***ich*** *want x* 'I

want x', ***ich kann nicht*** *x* 'I cannot x'). Essentially this feature of usage-based approaches suggests that the lexicon cannot be neatly categorized into discrete word classes. Rather, functional processes such as chunking, entrenchment and frequency-based activation might explain the child's mixing patterns (e.g. Backus, 2015). In many approaches, functional elements are thought to provide the grammar with some content words from the other language mixed into the frame (e.g. Myers-Scotton, 1997). However, as we have tried to show in our analyses, this asymmetry is difficult to maintain if functional elements are part of partially schematic units. Analyses of schemas showed that the child's main mixing consisted of a mixed schema (e.g. ***ich*** *want (x)*) and the open slot either unfilled or with English mixed in. Interestingly, if the schema was German, the child exclusively mixed English into the open slot. This could suggest that the fixed part of the utterance may, as suggested above, have come from routine interactions in his kindergarten and the open slot was filled by the context language that he was currently involved with.

It seems that Tim's mixed utterances are suggestive of a number of different 'production routes': Some utterances indicate the possibility of self-entrenched schemas containing both English and German. This could be similar to the self-entrenchment of errors in monolingual children (for instance children who use *me* or *my* as subjects for relatively extended period (Kirjavainen, Theakston & Lieven, 2009). The idea is that Tim may have preferred the use of ***ich*** to *I* and started to use it frequently with English *want* in the accepting home context and this usage then became entrenched as an 'easy to produce' string.

A series starting with *When* could be explained in a similar way. *When* is a subordinating conjunction and Tim used the correct verb-final pattern for German.

However, note how close the following utterances are, in both meaning and form, and that the conjunction is in English while the verb is German (*When I a little baby* ***ist*** 'When I a little baby is', *When I a little girl* ***ist*** 'When I a little girl is', *Tim when a baby* ***ist*** 'Tim when a baby is', *When Tim little boy* ***ist*** 'When Tim little boy is', *When I little, little* ***ist*** 'When I little, little is'). Another example of self-entrenched schemas shows an OVS pattern (for instance ***Ein*** *big one wants Tim* 'A big one wants Tim', ***Ein*** *yellow spoon* ***hat*** *Tim* 'A yellow spoon has Tim'). These are all with either ***hat*** 'has' or *want(s)* as the verb and, with one exception, Tim as the subject. About 20% of transitives in German Child Directed Speech (CDS) have this word order (Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008) but it is extremely unusual in English and, we would guess, almost non-existent in English CDS. It is possible that at this point in Tim's development, these had become somewhat entrenched schemas with open slots which were mainly filled with noun phrases - but note that these schemas show an emerging sensitivity to German word order rules.

What we seem to be seeing here is a fascinating interaction between the entrenchment of self-developed schemas, developing grammatical knowledge in both languages and creativity in utterance production. Vihman (1998) is right to point out that many of the mixed utterances of bilingual children are creative in that they are unlikely ever to have been heard and therefore directly learned from the input. This is certainly true of Tim's mixed utterances. One can see this fairly clearly if we look at some of the errors that Tim made. For instance, in Tim's utterances with negation there are examples of the correct negator form (***Ich keine*** *little munchkin* 'I no little munchkin', ***Du*** *is not mad* 'you is not mad') and correct placement of the negator in both German and English (***Ich kann nicht*** ... 'I can not' …*But there's no ketchup* ***mehr***

'but there's no ketchup more'). However, errors are equally or more common both with the form and placement (*No* ***keine*** *go at the doctor* 'no, no one go at the doctor'*,* ***Ich brauche ein nicht*** *bib* 'I need a not bib', ***Du nicht*** *cry* 'you not cry', ***Ich*** *go at the school no* 'I go at the school no'). In two of these utterances, *no* is followed by ***keine*** 'no/not/none', in both cases incorrectly preceding a VP. There are 3 utterances with ***keine*** incorrectly serving as a general negator (***Keine*** *out,* ***Keine*** *Mama friend,* ***Ich keine*** *go play)* possibly substituting for the use of English *not* which is often used by children learning English during the development of syntactic negation (Cameron-Faulkner, Lieven, & Theakston, 2007). It would be difficult to find just one explanation for these utterances. Much more likely is that some are generated directly from highly frequent strings in the input e.g. ***Ich kann nicht*** *x*, which our results find is the 11th most frequent schema used in Tim's code-mixes (see Figure 4). Others clearly originate with Tim's own creative productions but have become self-entrenched e.g. ***Du nicht*** *x* 'you not x', ***Ich nicht*** *x* 'I not x', ***Ich*** *not cried* 'I not cried'. Strings derived directly from the input are likely to be grammatically correct while, as we can see, self-entrenched schemas may well not be. Due to competition and reanalysis, these ungrammatical schemas will become less and less frequent, reflecting the fact that fully correct negation takes time to develop for some English-speaking children (Cameron-Faulkner et al., 2007) and some German-speaking children (Clahsen, 1983; Roeper, 1996).

The fact that Tim's mean length of utterance was longest for his mixed utterances has potential implications for understanding some of the processes underlying his production. Firstly, if he had entrenched partially schematic units which acted as chunks, then this would automatically lengthen his utterances. In addition, as just noted, if the schematic part was monolingual, it was likely to be grammatically

correct. On the other hand, there do seem to be a number of utterances where Tim's desire to get something said outstripped his current grammatical knowledge or, at least, this did not act as a constraint on his production. This is best exemplified by his errors of subject-verb agreement. Many utterances lacked any agreement, reflecting the 'optional infinitive' stage that monolingual children learning both English and German go through (Freudenthal, Pine, Aguado-Orea, & Gobet, 2007; Legate & Yang, 2007). However, there were also examples of errors of commission in agreement in Tim's mixed utterances: the use of 3rd person German ***hat*** 'has' with 1st and 2nd person noted above; 3rd person ***holt*** 'gets' with 1st person subject (from ***holen*** 'to get', ****Ich holt eine** pillow* 'I gets a pillow'); 3rd person English marking (****Ich** misses her Mama* 'I misses her Mama') and 2nd person German marking (**I **kannst nicht*** 'I can not') both with 1st person subject. Here the 'syntactic glue' seems to be absent. These errors were developmental in that Tim eventually stopped making them as both languages developed – even in mixed utterances, perhaps suggesting that it was the increasingly encapsulated representation of the two languages and the development of their grammars that 'drove out' these types of errors, but also functional mixes more generally.

Further, our analysis showed that if constructions are taken as the basic elements, the concept of a 'matrix language' acting as the 'syntactic glue' by which the grammatical frame can only come from one language is difficult to maintain. First, usage-based models do not assume a full lexicon-syntax divide and posit constructions as their basic linguistic units, which blurs the line between syntax and lexicon (e.g. Tomasello, 2003). Second, both functional and lexical elements can be part of a unit and/or partially schematic unit, and thus we cannot classify words into discrete word

classes. Usage-based accounts do not predict universal switching points in a sentence or which linguistic element can or cannot be switched. Rather, code-mixing is influenced by levels of entrenchment and abstractness. The suggestion is that many of the code-mixed utterances of bilinguals do not result from the mixing of individual functional and lexical elements but from constructions which are processed as one entity. Constructions and their level of abstractness are shaped by various factors such as frequency of input and usage. This is what we found for Tim: his code-mixing consisted in large part of partially schematic units such as ***ich*** *x it* 'I x it' with the open slot filled by either German and/or English elements, e.g. ***ich*** *zip it* 'I zip it', ***ich spielen*** *it* 'I play it'. Consequently, language (and grammar) are thought to be generalizations over usage. Usage is different for each speaker and can even result in self-entrenched, bilingual constructions which are not part of either language, e.g. ***ich*** *want* 'I want', ***und*** *this* 'and this'. These bilingual constructions may ease processing in the bilingual since they are processed as a whole unit.

In conclusion, the aim of the present study was to investigate the types of code-mixing made by a 2-year-old English-German-speaking child. We have suggested that an important basis for code-mixing is the use of partially schematic constructions. What on the surface can be seen and analyzed as code-mixing is often the use of a partially schematic construction from one language and the open slot filled with material from one and/or the other language. Based on this assumption, code-mixing can then be shown to result from the selection (i.e. either conscious selection or unconscious entrenchment-based activation) of fully specific chunks and/or partially schematic units

which can even result in entrenched, partially schematic, bilingual constructions in which code-mixing has become conventionalized.

References

Backus, A. (2015). A usage-based approach to code-switching: The need for reconciling structure and function. In G. Stell and K. Yakpo (Eds.), *Code-switching Between Structural and Sociolinguistic Perspectives.* Berlin: Mouton de Gruyter.

Bernardini, P., & Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: The Ivy Hypothesis. *Bilingualism: Language and Cognition, 7*, 49-69. doi: 10.1017/S1366728904001270

Bowerman, M. (1988). The 'no negative evidence' problem: how do children avoid constructing an overly general grammar? In J. A. Hawkins (ed.), *Explaining Language Universals.* Oxford: Blackwell.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Cameron-Faulkner, T., Lieven, E., & Theakston, A. (2007). What part of *no* do children not understand? A usage-based account of multiword negation. *Journal of Child Language, 34*, 251-282. doi:10.1017/S0305000906007884

Cantone, K. F. (2007). *Code-switching in bilingual children*. Dordrecht: Springer.

Cantone, K. F. & Müller, F. (2008). Un Nase or una Nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*, 118 (6), 810–826. doi:10.1016/j.lingua.2007.05.007

Clahsen, H. (1983). Some more remarks on the acquisition of German negation. *Journal of Child Language, 10*(02), 465-469.

Comeau, L., Genesee, F., & Lapaquette, L. (2003). The Modeling Hypothesis and child bilingual code-mixing. *The International Journal of Bilingualism, 7*(2), 113-126. doi: 10.1177/13670069030070020101

Culicover, P. W. & Jackendoff, R. (2005). *Simpler Syntax.* Oxford: Oxford University Press,

De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge: Cambridge University Press.

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics, 28,* 411-424. doi: 10.1017.S0142716407070221

Demirçay, D., & Backus, A.M. (2014). Bilingual constructions: Reassessing the typology of code-switching. *Dutch Journal of Applied Linguistics*, *3*(1), 30-44. doi: 10.1075/dujal.3.1.02dem

Deuchar, M. (1999). Are function words non-language-specific in early bilingual two-word utterances? *Bilingualism: Language and Cognition, 2,* 23-34.

Deuchar, M., & Quay, S. (2000). *Bilingual acquisition: Theoretical Implications of a case Study*. Oxford: Oxford University Press.

Di Sciullo, A.-M., Muysken, P. & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22, 1–24.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development, 79*(4), 1152-1167.

Eichler, N. (2011). *Code-Switching bei bilingual aufwachsenden Kindern: Eine Analyse der gemischtsprachlichen Nominalphrasen unter besonderer Berücksichtigung des Genus.* Tübingen: Narr.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S. & Bates, E. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual.* Baltimore, MD: Brookes.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science, 31*, 311-341.

Gawlitzek-Maiwald, I. & Tracy, R. (1996). Bilingual bootstrapping. *Linguistics, 34*, 901-926.

Genesee, F., Boivin, I., & Nicoladis, E. (1996). Talking with strangers: a study of bilingual children's communicative competence. *Applied Psycholinguistics, 17*, 427-442.

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language, 22*, 611-631. doi: 10.1017/S0305000900009971

Goodz, N.S. (1989). Parental language mixing in bilingual families. *Journal of Infant Mental Health, 10*, 25-44. doi: 10.1002/1097-0355(198921)10:1<25::AID-IMHJ2280100104>3.0.CO;2-R

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(02), 131-149. doi: 10.1017/S136672899800025X

Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing.

Hoff, E. (2003). Causes and consequences of SES-related differences in parent-to-child speech. In M. H. Bornstein & R. H. Bradle (Eds), *Socioeconomic status, parenting, and child development* (pp. 147-160). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoff, E. & Place, S. (2012). Bilingual language learners. In S. Odom, E. Pungello & N. Gardener-Neblett (Eds), *Re-visioning the beginning: Developmental and health science contributions to infant/toddler programs for children and families living in poverty* (pp. 77-101). New York: Guilford Press.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics, 19*(4), 603-629. doi: 10.1017/S0142716400010389

Huttenlocher, J., Vasilyeva, M., Cymerman, E. & Levine, S. (2002). Language input at home and at school: Relation to child syntax. *Cognitive Psychology, 45*(3), 337-374.

Joshi, A. (1985). Processing sentences with intrasentential code switching. In D. R. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language parsing*. New York: Academic Press.

Kirjavainen, M., Theakston, A. L., & Lieven, E. (2009). Can input explain children's me-for-I errors? *Journal of Child Language, 36*(5), 1091. doi: 10.1017/S0305000909009350

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*, vol. 1. Stanford University Press.

Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language, 19*, 633-658. doi: 10.1017/S0305000900011600

Lanza, E. (2004). *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford: Oxford University Press.

Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition, 14*, 315-344.

Lieven, E. and Behrens, H. (2011). Dense sampling. In E. Hoff (Ed.), *Guide to Research Methods in Child Language* (S. 226-239). Oxford: Wiley-Blackwell.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language, 30*(02), 333-370. doi: 10.1017/S0305000903005592

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20*(3), 481-507. doi: 10.1515/COGL.2009.022

MacSwan, J. (1999). *A Minimalist Approach to Intrasentential Code Switching*. New York: Garland Publishing.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marchman, V. A. & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9-F16. doi: 10.1111/j.1467-7687.2008.00671.x

Meisel, J. M., Ed. (1990). *Two first languages. Early grammatical development in bilingual children*. Dordrecht: Floris.

Meisel, J. M., Ed. (1994). *Bilingual first language acquisition. French and German grammatical development*. Amsterdam: Benjamins.

Müller, N., L. Arnaus Gil, N. Eichler, J. Geveler, M. Hager, V. Jansen, M. Patuto, V. Repetto & A. Schmeißer (2015). *Code-Switching: Spanisch, Italienisch, Französisch. Eine Einführung.* Tübingen: Narr.

Myers-Scotton, C. (1997). *Duelling languages: Grammatical structure in code-switching*. Oxford: Oxford University Press.

Myers-Scotton, C., & Jake, J. L. (2001). Explaining aspects of code -switching and their implications. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing*. Massachusetts: Blackwell Publishers.

Paradis, J. (2000). Beyond 'one system or two': degrees of separation between the languages of French-English bilingual children. In S. Döpke (Ed.), *Crosslinguistic structures in simultaneous bilingualism* (pp. 175-200). Amsterdam: John Benjamins.

Paradis, J., Nicoladis, E., & Genesee, F. (2000). Early emergence of structural constraints on code-mixing: evidence from French - English bilingual children. *Bilingualism: Language and Cognition, 3*(245-261).

Pearson, B. Z., Fernandez, S. C., Lewedeg, V. & Oller, D. K. (1997). 'The Relation of Input Factors to Lexical Learning by Bilingual Infants'. *Applied Psycholinguistics 18*(1): 41–58. doi: 10.1017/S0142716400009863

Petersen, J. (1988). Word-internal code-switching constraints in a bilingual child's grammar. *Linguistics, 26*, 479-493.

Petitto, L. A., Katerelos, M., Levy, B., Gauna, K., Tétrault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for mechanisms underlying bilingual language acquisition. *Journal of Child Language, 28*(2), 1-44.

Pfaff, C. W. (1979). Constraints on language mixing: Intrasentential codeswitching and borrowing in Spanish/English. *Language, 55*, 219-318.

Poplack, S. (1980). 'Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPANOL': Toward a typology of code-switching. *Linguistics 18*, 581-618.

Quick, Endesfelder A., Lieven, E., & Tomasello, M. (2016). Mixed NPs in German-English and German-Russian bilingual children. In Robinson, Justyna A. and Monika Reif (Eds), *Cognitive perspectives on bilingualism.* Trends in Applied Linguistics (TAL). Berlin: de Gruyter.

Redlinger, W. E. & Park, T. (1980). Language mixing in young bilinguals. *Journal of Child Language, 7*(2), 337-352. doi: 10.1017/S030500090000266X

Roeper, T. (1996). The role of merger theory and formal features in acquisition. In H. Clahsen (Ed.), *Generative perspectives on language acquisition* (pp. 415-449). Amsterdam: John Benjamins.

Rowland, C. (2007). Explaining errors in children's questions. *Cognition*, 104 (1), 106-34. doi: 10.1016/j.cognition.2006.05.011

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31*, 101-121. doi: 10.1017/S0305000903005944

Vihman, M. (1985). Language differentiation by the bilingual infant. *Journal of Child Language, 12*, 297-324. doi: 10.1017/S0305000900006450

Vihman, M. (1998). A developmental perspective on code-switching: Conversations between a pair of bilingual siblings. *The International Journal of Bilingualism, 2*(1), 45-84. doi: 10.1177/136700699800200103

Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry*, 14, 520–536.

Footnotes

[1]Upper Bound refers to the longest utterance of a child found in a recording.

[2]Throughout the paper German material is bolded.

Table 1

Coding for construction types

Code	Construction	Examples
NP	noun phrase	*the dog* *this one* ***keine*** *clothes* 'no clothes' *no a dog.*
AP	adjective phrase	***ein*** *big tiger* 'a big tiger' ***ein*** *big one* 'a big one' *and* ***ein*** *blue bed* 'and a blue bed'
PP	prepositional phrase	*on the couch*
VP	utterances with a verb, which is neither a copula verb, nor a complex verb	***und*** *wear* ***die*** *shoes* 'and wear the shoes'
Cop	form of *be* as a main verb	*this is* ***eine*** *tiger* 'this is a tiger' ***ich bin auch*** *sick* 'I am also sick' ***und*** *this is* ***ein*** *brown dog* 'and this is a brown dog'
Complex VP	sentences with complex verb such as modal verbs, tense forming	*Jona* ***hatte*** *fixed it* 'Jona had it fixed' ***ich*** *want* ***malen*** 'I want (to) paint'
SS	Simple sentence with main verb	*I want carrots* ***ich*** *feed birds* 'I feed birds' ***ich*** *fix roof* 'I fix (the) roof'
Complex S	Complex Sentence, sentences containing two verbs; basically subordinated or conjoined clauses	*I don't know* ***ich bin fertig*** 'I don't know I am done' ***Und des is mamas und des is papas*** 'and this is mom's and that is daddy's
Imp	Imperatives	*Help me.*

		*look **eine** bridge* ‘look, a bridge’
Quest	Question	***Und** this?* ’And this?’
		***Ich** hold this?* ‘I hold this?’
		What is that?
Frag	all utterances which do not fit into any other category	***du** baby tiger* ‘you baby tiger’
		*and **ja*** ‘and yes’

Table 2

MLU for monolingual and mixed utterances

	Recordings		Diary (2;6-2;7)	
	MLU	UB5	MLU	UB5
German monolingual	2.8	7.8		
English monolingual	3.6	12.4		
Mixed	4.1	13.4	4.4	8.2
All utterances	3.7			

Table 3

The percentage of constructions in the recordings and code-mixed data

Percentages	Recordings, English Monolingual	Recordings, German Monolingual	Recordings, Code-mixing	Diary, Code-mixing
Noun Phrase	7	14	7	5
Adjective phrase	2	1	2	2
Prepositional phrase	4	1	2	1
Verb phrase	6	7	2	4
Copula	15	6	9	4
Complex verb phrase	12	2	8	16
Simple sentence	13	15	28	43
Complex sentence	2	0	4	4
Imperative	11	10	5	3
Question	15	14	15	12
Fragment	13	30	18	6
Total number	3785	483	1331	277

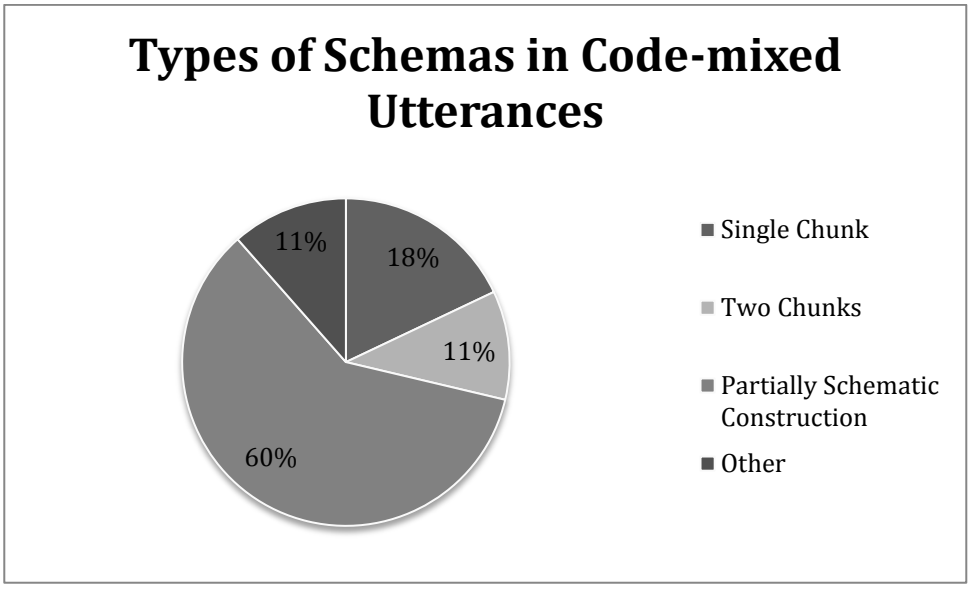

Figure 1. Distribution of different schema types in code-mixed utterances

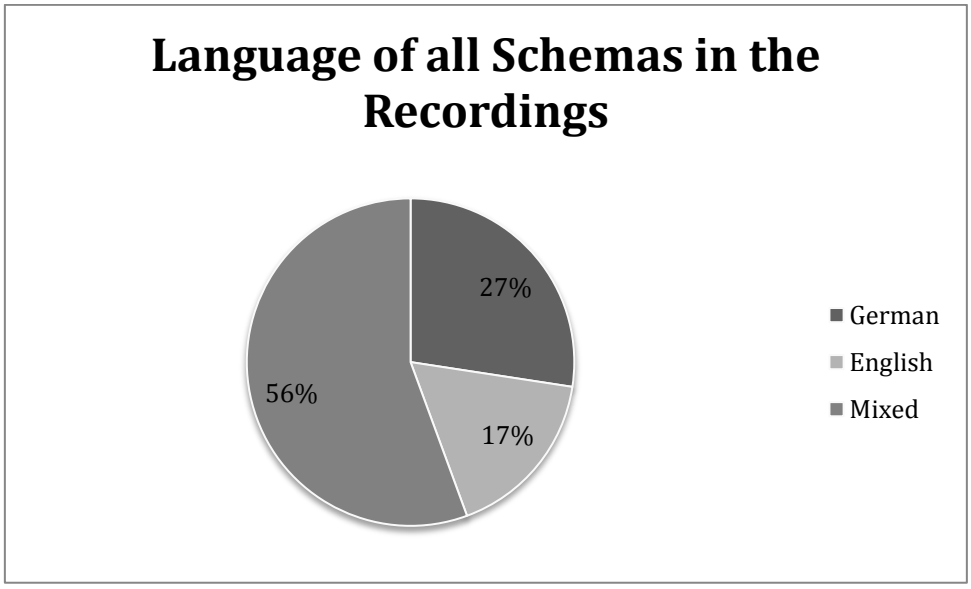

Figure 2. Language distribution in observed schemas in the recordings

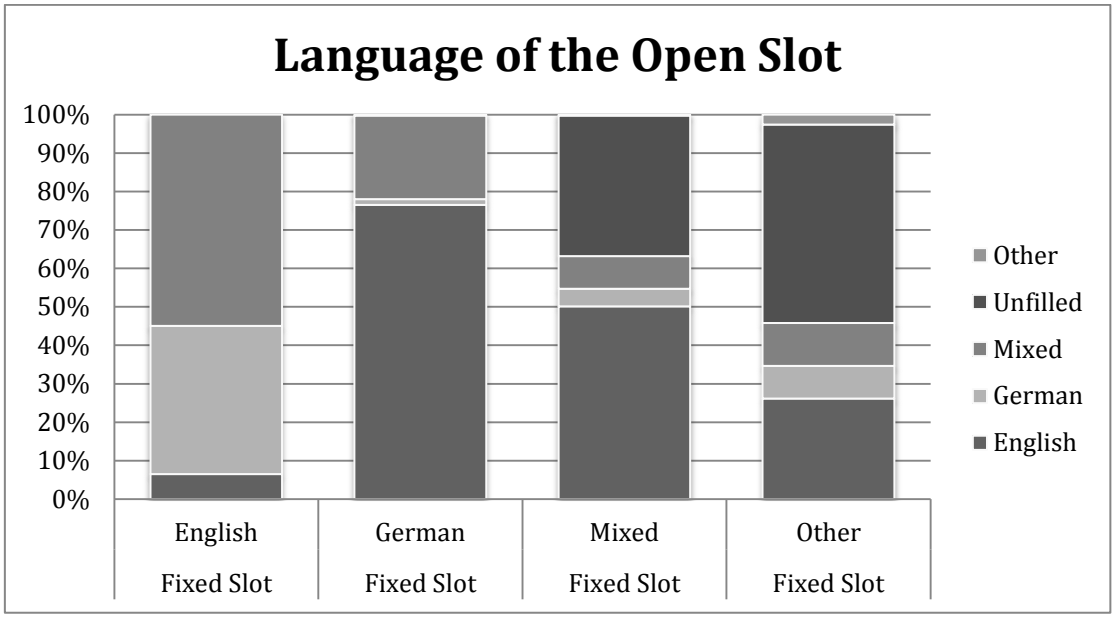

*Figure 3***.** Language distribution in the open slot

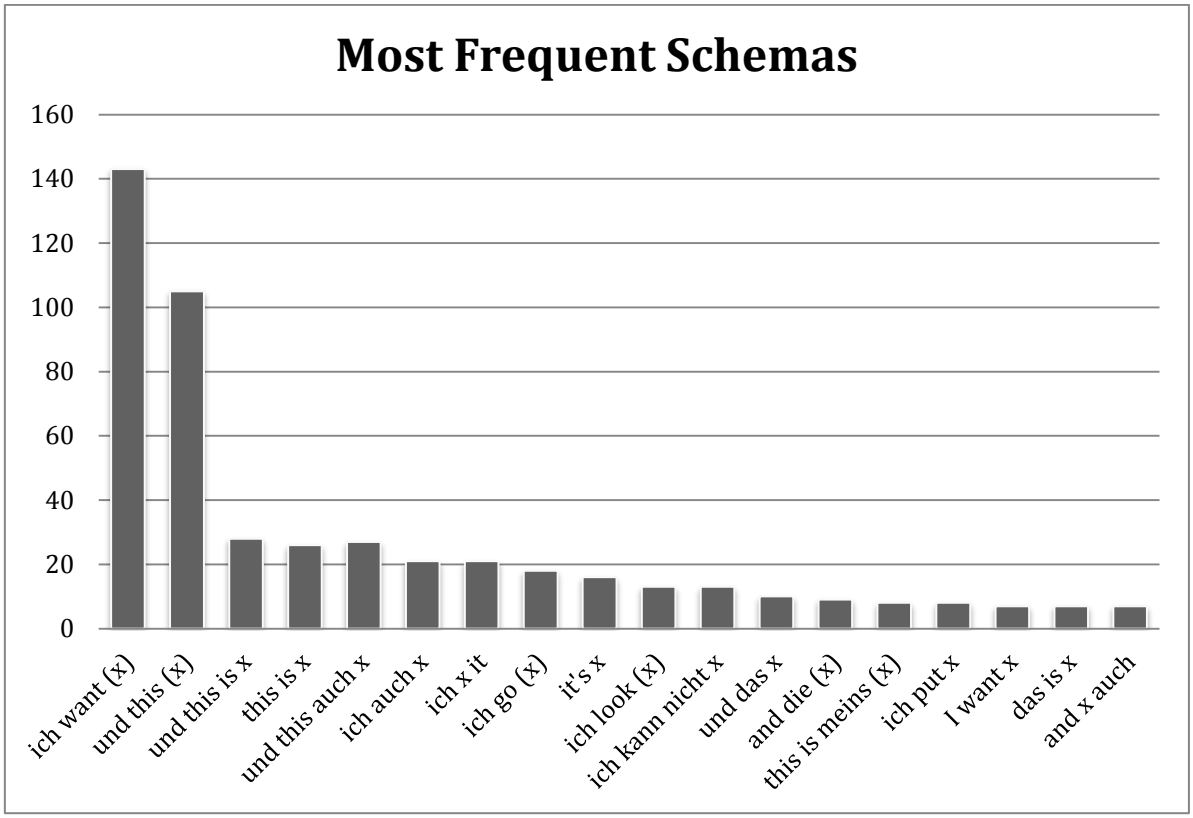

Figure 4. Number of most frequent schemas found in the code-mixed utterances