

Stefan Hartmann*, Nikolas Koch and Antje Endesfelder Quick

Cognitive Linguistics meets multilingual language acquisition: What pattern identification can tell us

<https://doi.org/10.1515/gcla-2022-0011>

Abstract: The usage-based approach to first language acquisition has become highly influential in research on first language acquisition. In recent research, it has also been adapted to account for language contact phenomena in multilingual first language acquisition, i.e. in situations in which children acquire two or more languages simultaneously. In this paper, we give a brief overview over these developments, summarize some first major results of this research program, and discuss remaining open questions and challenges. In particular, we review a number of studies that have used the traceback method, previously established in research on monolingual acquisition, to identify recurrent patterns in the early language of multilingual children, especially in their code-mixing, i.e. the use of more than one language in one utterance. We argue that the usage-based approach can help to shed light on some of the open questions in research in multilingual acquisition, especially as it is highly compatible with other prominent concepts in current research on multilingualism, and that it provides us with the methodological toolkit that is needed to investigate language contact phenomena in a data-driven way.

Keywords: multilingual first language acquisition; usage-based approach; trace-back method; code-mixing

1 Introduction

The question of how humans learn language(s) is a major cornerstone of the cognitive-linguistic enterprise – in fact, it has been one of the crucial questions of Cognitive Linguistics from early on. While nativist approaches posit an innate

* **Corresponding author: Stefan Hartmann**, Heinrich-Heine-Universität Düsseldorf, hartmast@hhu.de

Nikolas Koch, LMU Munich, koch@daf.lmu.de

Antje Endesfelder Quick, Universität Leipzig, antje.quick@uni-leipzig.de

Universal Grammar, the approaches that can be summarized under the umbrella term “cognitive linguistics” assume that children learn their language(s) from the input they receive on the basis of domain-general cognitive capacities such as pattern finding and intention reading (Tomasello 2009).

In particular, the approach put forward most prominently by e.g. Michael Tomasello and Elena Lieven (e.g. Tomasello 2003; Tomasello and Lieven 2008), often dubbed “the usage-based approach to language acquisition”, has become highly influential in Cognitive Linguistics and beyond. This paradigm assumes that children start learning their first language in a piecemeal way by acquiring fixed chunks (holophrases), followed by so-called pivot schemas organized around concrete words (e.g. *more milk*, *more juice*), and item-based constructions, i.e. frame-and-slot patterns like [*This is X*]. A number of studies have shown the item-based nature in the output of child language in corpora and experimental data (e.g. Bannard and Matthews 2008; Lieven et al. 2009) as well as in child-directed speech (e.g. Cameron-Faulkner et al. 2003).

The usage-based approach to first language acquisition has largely focused on monolingual acquisition, and mostly on children that learn English as their first language. From a global perspective, however, monolingual acquisition is the exception, not the norm: Crystal (2003: 17), for instance, estimates that around two-thirds of all children grow up in multilingual environments and are hence likely to learn more than one language simultaneously. As a general tendency, much research on multilingual acquisition has largely taken a more formalist perspective, even though there are some exceptions (e.g. Vihman [1999], who adopts major theoretical and methodological aspects of the usage-based paradigm). Over the last decades, however, and especially in recent years, there have been first steps towards applying the usage-based theory of language acquisition to multilingual acquisition (see Backus [2020], for an overview). From a usage-based perspective, there is no reason to assume a principled difference between monolingual and multilingual acquisition. However, there are some phenomena that are unique to multilingual language acquisition and use which potentially present a challenge – and as such arguably also an ideal testing ground – for usage-based theory. Such potentially challenging language contact phenomena include structural transfer from one language to another, e.g. adopting syntactic structures from one language in another language as in (1) or, perhaps most prominently, code-switching or code-mixing, i.e. the use of more than one language in a single speech event as in (2).¹

¹ In this paper, we focus on the acquisition of spoken language(s). On code-switching between signed languages, see e.g. Quinto-Pozos (2009).

- (1) Engl. *we have gone to school in Tarrington* → German *Wir haben gegangen zu Schule in Tarrington* (instead of *Wir sind in Tarrington zur Schule gegangen*) (from Clyne 2003: 79)²
- (2) *look at the Ampel, it's kaputt* 'look at the traffic light, it's broken' (Fion, 03;10.22)³

These phenomena can offer new insights to established usage-based accounts of language acquisition. Children acquiring two languages often acquire them in wildly different circumstances. Some children grow up in bilingual families, others in societies in which multilingualism is the norm and yet other children grow up in monolingual families living in an L2 society. From a usage-based perspective, the question arises how these differences influence the acquisition process, whether the linguistic knowledge varies considering these many different input situations, and finally also, how dynamic the output mirrors the input situations. If we assume that children acquire constructions by abstracting away patterns from the input they receive, a usage-based account should be able to accommodate the various differences in the way children acquire their languages, and we should be able to explain how patterns feature in contact phenomena.

In this paper, we address the question of how the usage-based approach deals with language contact phenomena in first language acquisition, how it may re-perspectivize some of the key questions that are frequently discussed in research on multilingual first language acquisition,⁴ and which methodological approaches can be used to study multilingual acquisition from a usage-based perspective. In Section 2, we will first give a brief overview of the usage-based approach and its implications for multilingual acquisition. In Section 3, we review a series of case studies that have used quantitative-explorative approaches like the traceback method (Hartmann et al. 2021) to study code-mixing in early bilingual speech. In Section 4, we summarize the main results that can, in our view, be drawn from these studies, but we also point to remaining open questions.

² Note that this example actually comes from an adult speaker as Clyne's study does not focus on first language acquisition; it is used for illustration only.

³ See Section 3 for a more detailed description of the corpora cited in this and the subsequent examples. As usual in first language acquisition studies, we refer to the age of a child in the format "years;months.days", i.e. 03;10.22 means that Fion was 3 years, 10 months and 22 days old when he made this utterance.

⁴ In the present paper, we are only concerned with simultaneous multilingual L1 acquisition, leaving out other acquisition contexts such as L2 acquisition in early childhood.

2 Monolingual and multilingual acquisition from a usage-based perspective

Tomasello (2009: 69) summarizes the usage-based approach to linguistic communication in two aphorisms: “[m]eaning is use”, and “structure emerges from use”. For language acquisition, this entails that children learn all aspects of their language(s) – grammatical structure, meaning, as well as pragmatic dimensions of language use (e.g. Bruner 1983: 18) – from the input they receive. These ideas are close in spirit to the framework of linguistic emergentism, which shares many ideas with usage-based theory and has also been highly influential in research on child language acquisition (see e.g. MacWhinney [2015] for an overview of emergentism).

Usage-based theory has been supported by a vast amount of empirical evidence (see e.g. Ambridge and Lieven [2011]; Lieven [2014] for overviews). According to usage-based theory, children first learn individual words and so-called holophrases, i.e. recurrent fixed chunks such as *What’s that?*, followed by word combinations, pivot schemas, and item-based constructions, as shown in Figure 1 (e.g. Tomasello 2003: 31–40; Tomasello 2009). Word combinations are simple juxtapositions of words (or holophrases), while pivot schemas show a more systematic pattern (Tomasello 2009: 76). The term “pivot” goes back to Braine (1963) and refers to constructions in which linguistic material is organized around specific words, usually verb-like predicative words (see Tomasello 1992: 20). For instance, children often use [*more* X] and [X *more*] interchangeably in their early speech, indicating that *more* serves as a pivot here. Pivot schemas can be considered early (constructional) patterns, but importantly, they do not have syntax (Tomasello 2003: 115). According to Tomasello (2003: 155), “[t]he consistent ordering patterns in many pivot schemas are very likely direct reproductions of the ordering patterns children have heard most often in adult speech, with no communicative significance.” The last stage are item-based constructions, which differ from pivot schemas in that children use, for instance, morphology, adpositions, and word order to syntactically mark which roles the different participants play in the events described (Tomasello 2003: 120).



Figure 1: Four stages of construction acquisition according to Tomasello (2003).

The development from early word combinations through pivot schemas to item-based constructions which are used ever more productively over time has been investigated using a broad variety of different methodological approaches. One of them, the traceback method, will be introduced in more detail in Section 3. Most of those studies were conducted on monolingual material. From a usage-based perspective, however, we can expect that this account also holds for multilingual acquisition. Thus, it seems highly promising to try and identify early “pivots” as well as item-based constructions in multilingual data. Importantly, there is no reason to assume that the two or more languages being acquired by the multilingual learner form strictly distinct repertoires. As Grosjean (1989: 3) put it, “[t]he bilingual is not two monolinguals in one person.” If we assume that language acquisition is item-based, what matters are the linguistic items that a learner perceives and produces, no matter which language they are coming from.

This is very much in line with other approaches that have become highly influential in research on language contact and multilingualism. To mention only two of those here, the concept of “linguistic multi-competence” (e.g. Cook 2016) emphasizes language users’ capability to flexibly make use of their full linguistic repertoire in specific communicative situations. In a similar vein, the concept of “translanguaging” questions the classic construal of “languages” as (more or less) closed systems, and instead treats different languages primarily as socio-cultural constructs (Otheguy et al. 2015). This does not mean, however, that constructions from different languages mash together to an undifferentiated mass. Instead, following Goldberg’s (2019: 17) definition of constructions as “emergent clusters of lossy memory traces that are aligned within our high- (hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions”, and given the observation that language users’ knowledge about virtually every single linguistic unit they know is astonishingly rich (e.g. Taylor 2012: 282), it seems reasonable to assume that multilingual learners gradually come to associate words and constructions from different languages with different situations and usage contexts, as well as with different collocation and frequency profiles. As the child’s metalinguistic awareness grows, different languages can also come to be associated with different interactional or emotional aspects. This metalinguistic awareness becomes particularly obvious in children’s requests to switch the language, as exemplified in (3).⁵

- (3) a. nein deutsch sprechen. ‘nein German speak’ (Fion, 03;05.11)
 b. you are speaking deutsch. ‘you are speaking German’ (Fion, 03;05.16)
 c. komm speak in deutsch. ‘come on, speak in German’ (Fion, 03;05.17)

⁵ See Section 3 for a description of the corpora cited in this and the subsequent examples.

One of the possible associations that the child could develop is which of their parents is predominantly using items from each language. Especially in Western cultures, parents often deliberately keep their languages apart (Gardner-Chloros 2009: 144), thus following a “One-Parent-One-Language” (OPOL) approach (Barron-Hauwaert 2004). Example (4) provides an insight into how the OPOL strategy is negotiated between child and parent in actual practice.

- (4) FAT: salat. ‘(German) *Salat*.’
 CHI: it’s lettuce!
 FAT: bei mir bei mir is das salat silvie. ‘For me, this is *Salat*, Silvie.’
 CHI: bei mammi is das lettuce. ‘For Mommy, this is *lettuce*.’
 FAT: nun guck mal! guck wenn wir beide zusammen reden dann kannst dann reden wir deutsch klar ja? wenn du. ‘Now look. When the two of us are talking, then we speak German, okay? When you...’
 CHI: salat!
 FAT: silvie?
 CHI: mhm.
 FAT: silvie? wenn wir beide zusammen reden dann redest du und ich wir reden deutsch klar? ‘Silvie? When the two of us are talking, the you and I, we speak German, okay?’
 CHI: reden deutsch. ‘Speak German.’
 FAT: und mit mammi sprichst du englisch! ‘And with Mommy, you speak English.’
 CHI: englisch. ‘English.’ (Silvie, 02;05.04)

The investigation of multilingual language acquisition is thus intimately connected to many of the key tenets of usage-based and cognitive-linguistic research. Firstly, we want to understand the cognitive principles behind language acquisition in general and multilingual acquisition in particular, which is in line with the “cognitive commitment” of cognitive linguistics as posited by Lakoff (1990). Secondly, the intimate connection between language and social, cultural, as well as interactional factors which has been taken into account more and more seriously in cognitive linguistics (see e.g. Croft [2009] for an early programmatic paper) is also evident in multilingual acquisition, as the examples just discussed have shown. Thirdly, research in cognitive linguistics has come to emphasize individual differences between speakers (e.g. Dąbrowska 2012, Quick et al. 2021b). Multilingualism is a particularly interesting field of exploration for investigating individual differences, as there are a number of parameters that influence multilingual speakers’ use of their languages (e.g. the degree to which they are exposed to each language; the socio-cultural environment, etc.).

These aspects have of course been taken into account in research on multilingual acquisition, regardless of whether they were rooted in the usage-based paradigm or not. Cognitive and constructionist approaches arguably provide a unified framework for understanding those different aspects. They assume that form-meaning pairs (constructions) are learned in specific contexts and can be understood as rich association patterns which do not only comprise phonological and semantic information but also pragmatic and discourse-functional properties (e.g. Croft 2001: 18). As Schmid (2020: 343) puts it, “Not only linguistic knowledge in a narrow sense, i.e. phonological, morphological, lexical, and grammatical knowledge, but also cotextual, situational, interpersonal, and social aspects of usage are entrenched as parts of routinized patterns of associations”.

In the remainder of this paper, we will focus on approaches that aim at uncovering the “building blocks” of multilingual language acquisition on the form side. Those, however, can also give us first clues to functional aspects as well as to individual differences between speakers; a more thorough understanding of multilingual acquisition, however, would of course require a systematic triangulation of empirical evidence from different sources. We will briefly touch upon this desideratum in the final section of this paper.

3 Case studies

3.1 Detecting chunks and frame-and-slot patterns in (bilingual) data

In a series of recent studies (e.g. Quick et al. 2021a, Quick and Hartmann 2021), we have applied one particular method originally developed for investigating monolingual acquisition to corpora that document bilingual language acquisition, namely the traceback method (e.g. Dąbrowska and Lieven 2005). In the traditional application of this method, a longitudinal corpus documenting the language acquisition of one child is split into two parts: The so-called test corpus, which usually contains the last two recording sessions, and the main corpus, which contains all previous recordings. The goal of the method is to show that the vast majority of the child’s utterances in the test corpus have predecessors in the main corpus, i.e. they are either verbatim repetitions (so-called “fixed strings”), or they can be accounted for with the help of “templates” that are partially lexically specific and contain an open slot, such as [I want REFERENT], so-called frame-and-slot patterns. In order to detect frame-and-slot patterns, the traceback

method uses a limited number of operations. Different implementations of the method use different operations; typically, three operations are used: SUPERIMPOSE, SUBSTITUTE and ADD, exemplified in (5) with examples from Hummel (2006).

(5) SUBSTITUTE

I want my REF
banana → *I want my banana.*

SUPERIMPOSE

I want my REF
my banana → *I want my banana.*

ADD

I want my banana
+ now → *I want my banana now.*
 or: *Now I want my banana.*

In practice, a traceback analysis proceeds as follows: First, the analyst, or the algorithm, searches for verbatim precedents of each target utterance in the test corpus. If a match is found, the utterance counts as successfully derived, i.e. it is assumed that the child has it readily available. If no verbatim match is found, we search for partial matches, accounting for the target utterances with the help of the above-mentioned operations. If, for example, the target utterance is *I want my banana*, and the main corpus contains the utterances *I want my apple* and *I want my pear*, then we can postulate the frame-and-slot pattern [*I want my X*]. If, additionally, the item(s) that fill(s) the slot is attested in the main corpus – here: the lexical item *banana* –, then the utterance is seen as successfully derived. The motivation behind this approach is that if it can be shown that the later, usually more complex utterances can be accounted for with the help of fixed chunks and frame-and-slot patterns, this lends support to the usage-based account of language acquisition as sketched above.⁶

⁶ A reviewer asks the very legitimate question why the utterances are traced *back*, rather than taking the inverse approach of tracing patterns from early utterances to later ones. The reason for this is that children's utterances get more complex over time, so tracing later utterances back to previous ones arguably makes the method more conservative. However, there have also been adaptations of the method that take a traceforward approach (e.g. Bannard et al. 2009). Importantly, though, the goal of traceforward is slightly different from that of traceback in that it aims at providing a predictive grammar of children's utterances.

A number of studies employing the traceback method have shown that a large proportion of children's target utterances can be successfully derived from precedents in the main corpus (see Lieven et al. [2003]; Dąbrowska and Lieven [2005]; Lieven et al. [2009]; Vogt and Lieven [2010] for English data; see Miorelli [2017] for Italian; see Koch [2019] for German). These studies consistently find that a large proportion of utterances in the test corpus can be traced back to earlier utterances. This provides evidence in favor of the usage-based hypothesis that language acquisition is strongly item-based. In addition to testing this hypothesis, however, traceback has also been come to be used as an explorative method to identify patterns in children's speech (Hartmann et al. 2021; see Kol et al. [2014] and Koch et al. [2020] for critical evaluations of the traceback method).

In their application of the traceback method to bilingual acquisition, Quick et al. (2018a, 2018b, 2018c, 2021a, 2021b), Gaskins et al. (2021), and Quick and Hartmann (2021) have focused on code-mixing, i.e. the use of more than one language in one single utterance. As mentioned above, code-mixing might, at first glance, be seen as a challenge to constructionist approaches to language acquisition, as mixed utterances like *ich habe drei wheels* 'I have three wheels' (Fion, German-English, from Quick et al. [2018b]) seem to be examples of highly productive and creative, rather than formulaic, language use. On closer inspection, however, we may find that highly entrenched patterns also play an important role in bilingual utterances that show code-mixing. Just as early monolingual language acquisition is highly formulaic and characterized by low-scope schemas which seem to function as a gateway to greater complexity, a usage-based approach to multilingual acquisition would expect that code-mixed utterances arise from the combination of entrenched chunks from different languages and/or the productive use of formulaic patterns with one or more open slots. This is precisely what the studies reviewed below have set out to investigate.

3.2 Participants

In the following, we will first focus on studies investigating the acquisition of three German-English bilingual children. All of them grew up in one-parent-one-language households, lived in Germany, came from a middle-class household, and are simultaneous German-English bilinguals. The first child, "Fion", is the second child to a German-speaking mother and an English-speaking father. Although the parents mostly adhered to the OPOL strategy when they talked to Fion, they did not settle on a family language and sometimes used both languages interchangeably when all family members were present. Fion was recorded two

hours per week, yielding ca. 200 hours of recording covering the age range from 2;3 to 3;11.⁷

The second child, “Silvie”, had an English-speaking mother and a German-speaking father. The father’s proficiency in English was very limited and the parents therefore spoke German to one another. The corpus covers recordings from 2;4 until 3;10 years of age, 135 hours in total averaging to about 2.5 h of recordings per week.

Finally, “Tim” is the only child of an English-speaking mother and a bilingual Spanish-Catalan father. Technically, Tim grew up trilingually, with the mother speaking mostly English to him and the father mostly Spanish, while a babysitter with whom he spent ten hours per week from 0;5 onwards spoke German to him, and German was also the language of the kindergarten he attended for more than 30 hours per week from 1;3 onwards. Effectively, Tim was largely bilingual as his exposure to and use of Spanish was very limited, as English was used as the family language. Tim was recorded 4 to 6 times per week, covering a period of six weeks between the age of 2;6 and 2;7, yielding a total of 30 hours of recordings.

3.3 Traceback and multilingual data

In order to check whether code-mixed target utterances can still be accounted for with the help of component units from a different subcorpus, Quick et al. (2018c) carve up their data differently from the traditional traceback studies mentioned above: Studying Tim’s bilingual language use, they use the set of all code-mixed utterances as their test corpus, and trace each code-mixed target utterance to earlier monolingual and bilingual utterances. For example, the bilingual target utterance *this is kaputt* ‘this is broken’ contains the pattern [*this is* X], which is attested multiple times in earlier utterances, and the open slot is filled by a German element.

Among the patterns identified by the method are also bilingual units like *und this*, which the authors interpret as temporarily entrenched units. Of course this does not mean that the child only has these items available as fixed chunks (which would be odd, given that we can hardly expect them to occur in the input);

⁷ For comparison: Tomasello and Stahl (2004: 105) assume that a child is awake about ten hours a day, producing and comprehending linguistic structures during that time. Thus, recording two hours per week roughly covers 2.9% of the child’s waking hours. But this is of course only a very rough estimate, and we can expect considerable individual differences between children’s waking hours.

rather, the child has used them often enough that they have “grown together”, as it were.

Overall, the traceback of lexically fixed items was successful in 88% of the cases, which suggests that lexically restricted patterns play an important role in the production of code-mixed utterances. The authors conclude that the form and production of highly formulaic constructions in code-mixing closely follows the entrenchment of the children’s own patterns, presumably extracted from the input – but also repeated frequently by the children themselves. Table 1 shows the most frequent patterns identified in Tim’s code-mixed data, showing that some multilingual utterances can be accounted for with the help of patterns with monolingual fixed parts like *this is x*, while in other cases, bilingual word combinations are identified as patterns because they are so frequent in the data, e.g. *ich want x*.

Table 1: Most frequent patterns in Tim’s data, as identified by the traceback method.

Pattern	Frequency
<i>ich want x</i>	143
<i>und this (x)</i>	105
<i>und this is x</i>	28
<i>this is x</i>	26
<i>und this auch x</i>	27
<i>ich auch x</i>	21
<i>ich x it</i>	21
<i>ich go (x)</i>	18
<i>it’s x</i>	16
<i>ich look (x)</i>	13

In another study, Quick et al. (2021a) used the traceback method in a slightly different way. Again, all code-mixed utterances formed the test corpus, and they used a) the child’s monolingual utterances and b) the child’s input as their main corpus.⁸ This study, which investigates the language use of “Fion”, draws on an automatic analysis of utterance-initial uni-, bi-, tri-, and 4-grams. This means that an utterance like *und das da go to school* (Fion, 02;05.29; see Hartmann and Quick [2021]) is split into chunks of 1 word (*und, das, da...*), 2 words (*und das, das da,*

⁸ Strictly speaking, the term “traceback” may be a bit misleading here since the target utterances are not traced to previous recording sessions – however, we will keep the term here for convenience. Assuming that the input significantly shapes the child’s linguistic knowledge, one can also interpret the metaphor of “tracing back” utterances in a slightly different sense here.

da go...) etc. up to 4 words. For the n-gram analysis, only fixed chunks are taken into account, i.e. no slot-and-frame patterns are posited. This necessarily entails a lower traceback success than could be expected if frame-and-slot patterns were taken into account as well. To make the procedure even more conservative, the authors excluded primed utterances: Patterns that are attested in the 20 preceding utterances (regardless of the speaker, i.e. whether it is an utterance of the child himself or one of his caregivers) were seen as potential primes and therefore not taken into account. Also, the authors operationalize two different thresholds: Following the above-mentioned CDS studies by Cameron-Faulkner et al. (2003) and Stoll et al. (2009), they only consider n-grams to be (potentially entrenched) patterns if they occur at least four times in the test corpus, not counting the potential primes in the 20-utterance window just mentioned. In their analysis, they distinguish between patterns that occur at least twice in the main corpus and patterns that occur at least once in the main corpus.

Even with the most conservative parameter settings, a fairly large number of target utterances can be traced both to the input and to the child's monolingual utterances. Overall, 82.2% of Fion's utterances are identified as containing utterance-initial patterns, i.e. units that are attested at least four times in the test corpus. However, this number also contains one-word patterns, i.e. utterances in which only the first word is identical four or more times. If we limit the scope to multi-word units, the percentage is much lower, but still, more than half of all utterances (51.7%) can be accounted for. Zooming in on the multi-word units only, 63.6% of the identified patterns are attested at least twice in the monolingual data and 65.2% are attested at least twice in the input data (Quick et al. 2021a).

Given the fact that we are dealing with highly non-standard patterns like *und this* or *nein this is*, this can mean one of two things: Either the results lend strong support for the hypothesis that language is highly formulaic – or they show that the traceback method is actually too permissive and that it identifies patterns where it makes no sense to assume an entrenched pattern. There are at least two arguments that speak against the latter suspicion: First of all, the traceback method is characterized by an “inbuilt conservativeness” (Quick et al. 2018b: 298), and the operationalization of the method in this particular set of studies ensures that it is applied in an even more conservative way than in the case of the classic traceback studies. Secondly, the implementation of the method is very much data-driven in these studies, i.e. without positing any patterns in a top-down fashion.

Quick et al. (2018b) take an in-depth look at a sample of the Fion data. They manually identify frame-and-slot patterns to investigate whether Fion's code-mixed utterances are complete or partial repetitions of earlier utterances. As with the “Tim” data in Quick et al. (2018c) mentioned above, each code-mixed utterance was traced back to all utterances prior to the target utterance. Here,

however, the traceback results actually only constitute the starting point for a follow-up analysis in which they investigate how the identified frame-and-slot combinations develop: For instance, they show that the predominant language in the open slot changes over time from English to German (and vice versa for the fixed lexical items in the fixed part of the frame-and-slot pattern), e.g. *ich bin eine bee* ‘I am a bee’ (Fion, 02;07.17) – *and that ’s rueckwaerts* ‘and that’s backwards’ (Fion, 03;11.24). This is in line with Fion’s linguistic development in general, which includes a shift in his dominant language from German to English (see Figure 2) after a trip to Ireland and a visit from his monolingual English grandparents (Quick et al. 2018b: 289–290).

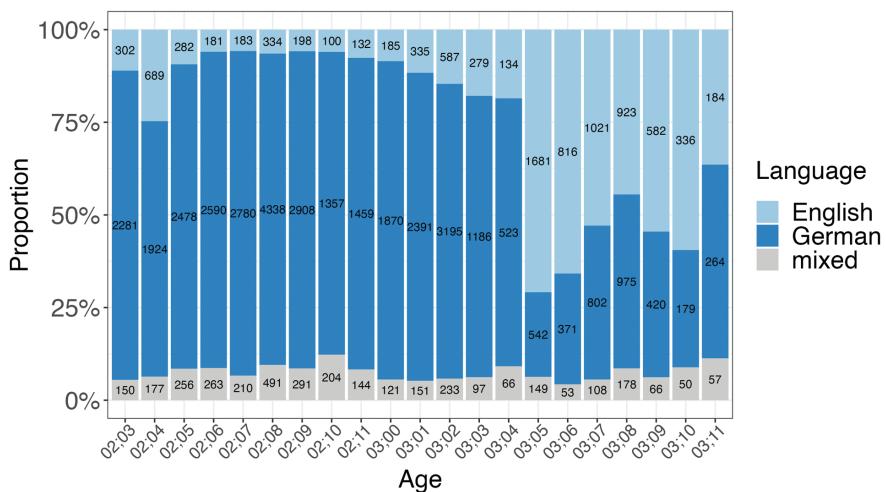


Figure 2: Language proportions in the “Fion” data, indicating a clear shift towards English around the age of 03;05.

Quick et al. (2018b) also go on to investigate discourse priming on the basis of their traceback results, showing that much of the code-mixing turns out to be primed by the occurrence of the same forms in the immediately preceding context, as in the two directly adjacent utterances in (6). In addition, they show that at an earlier age, the entire chunks tend to be primed, while at later age spans, only the open slots tend to be primed.

- (6) FAT: *I see a helicopter.*
 CHI: *I see a Kelle.* ‘I see a ladder’

This suggests that different input situations lead to considerable individual differences in the children’s output. In general, the role of individual differences is a crucial aspect that has to be taken into account when investigating child

language acquisition. It is therefore quite instructive to compare the traceback results obtained for different children. Quick and Hartmann (2021) do exactly this for Fion's and Silvie's data. They find that the traceback success differs considerably between the two children, as Fion's data contain much more code-mixed chunks that he re-uses multiple times. Code-mixing is thus quite a heterogeneous phenomenon that shows considerable individual differences. Even though it was not originally conceived for this purpose, the traceback method can arguably contribute valuable insights on this aspect.

In a similar vein, Gaskins et al. (2021) apply the traceback method to the Silvie data as well as to the data of two other bilingual children: "Sadie", a Polish-English bilingual, and "Eetu", a Finnish-English bilingual. Gaskins et al. (2021) use the children's code-mixed multi-word units as their test corpus and trace them back to the same children's monolingual constructions. Unlike Quick et al. (2021a), they manually identify frame-and-slot patterns, and they show that between 63 and 65% of all code-mixed target utterances can be traced back to the monolingual data and that across all children in more than 90% of all cases, monolingual frames are combined with slot fillers from the other language. They also show that, although the three children, in frame-and-slot patterns, have clear preferences for using fixed parts from one language and slot fillers from the other, there is also a fair share of frame-and-slot patterns in the data where the opposite direction can be observed: "17% of Sadie's frames were Polish; 4% of Sylvie's were English and 9% of Eetu's frames were Finnish" (Gaskins et al. 2021).

These examples show that the traceback method can provide interesting insights regarding bilingual language acquisition. The findings from Gaskins et al. (2021) and Quick and Hartmann (2021) also hint at various ways in which the method can be used to investigate individual differences. However, the method does have its limitations. It is, after all, based on arbitrary frequency thresholds, and it cannot take into account many of the more subtle aspects that play a role in the acquisition and use of constructional patterns. Thus, not all patterns detected by the method are cognitively plausible to the same extent. For instance, [*nein* X] is identified as a frame-and-slot pattern on a par with, say, [*this is* X], although it might be more straightforward to assume that *nein* is a "pivot" in the sense of Tomasello as explained above, which is flexibly combined with other linguistic material. Another limitation, at least in the way the method is implemented in the studies reviewed above, is that it operates at the level of words and (adjacent and non-adjacent, but always utterance-internal) word combinations. Thus, the results are only informative about constructional patterns at the syntactic level, although it would also be interesting to take morphological constructions into account as well, considering code-mixed utterances like *firewehr* 'fire brigade' (Fion, 02;05.24) or *I'm suching my Uhr* 'I'm looking for my watch' (Fion, 03;08.10).

The traceback results therefore have to be interpreted with some caution. Nevertheless, traceback provides one way to reach one of the central goals of usage-based approaches to language acquisition: It helps identify recurrent patterns and, by proxy, constructions. As such, it also paves the way for further analyses. We have discussed some examples above, but there are a number of other questions that could be pursued, particularly with regard to the frame-and-slot patterns. For instance, it would be interesting to investigate whether the slot fillers of a particular frame-and-slot pattern share certain formal features (e.g. their prosodic pattern), and whether items from the other language are more likely to occur in the open slot of frame-and-slot patterns in code-mixed utterances if they exhibit those features as well.

4 Conclusion

In this paper, we have given an overview of traceback approaches that aim at investigating multilingual first language acquisition from a usage-based perspective. While multilingual acquisition has been a major topic in language acquisition research for a fairly long time, and much research has actually implicitly shared many of the key assumptions of usage-based approaches, it was only recently that researchers have started to look at multilingualism from a decidedly usage-based, cognitive-linguistic, and constructionist perspective (for a survey, see Backus [2020]). A usage-based perspective entails a particular focus on the “building blocks” of language acquisition, i.e. the constructions that children learn. Uncovering those building blocks is a key tenet of usage-based approaches to language acquisition in general, and to multilingual acquisition in general. As an example for first steps in this direction, we have focused on our corpus-based studies using the traceback method, which has been established in usage-based research on monolingual acquisition and yields interesting results for bilingual data as well. We have summarized the main results of previous studies in which we have applied the traceback method to bilingual data, and we have discussed the potential and limitations of the method.

This opens up potential perspectives for future research, both regarding further steps in the exploration of the the datasets introduced here and regarding more general open questions that should be addressed in usage-based approaches to multilingual first language acquisition. As mentioned above, traceback focuses on the form side of constructions, even though the meaning side is taken into account indirectly as the open slots are often constrained semantically (via slot categories like PROCESS or THING, see e.g. Dąbrowska and Lieven [2005]).

Similarly, other pattern identification methods such as McCauley and Christiansen's (2017, 2019) Chunk-Based Learner or Ibbotson et al.'s (2019) Dynamic Network Model, which can both be seen as an extension of traceback, also focus on purely formal and distributional aspects. Thus, one open question is how semantic, but also e.g. phonological/prosodic aspects can be taken into account more systematically in characterizing patterns in early language acquisition.

As for multilingual language acquisition, one question that deserves to be investigated in more detail is to what extent code-mixing can be merely accounted for in terms of the filling of temporary lexical gaps and to what extent it may serve pragmatic and interactional purposes as well. In adult code-switching, it is a well-established finding that code-switching fulfills social functions, such as marking social identity (e.g. Gardner-Chloros 2009). As we have seen in Section 2, a considerable degree of metalinguistic awareness is already present in relatively early stages of language acquisition, which raises the question whether code-mixing may be used consciously in our data in some cases. Also, our studies reviewed above have solely focused on utterance-internal code-mixing. In order to get a fuller picture of how the children in our datasets make use of their languages, code-switching between utterances should be taken into account in future studies as well.

In sum, a cognitive-linguistic and usage-based perspective can add valuable insights to research on multilingual first language acquisition, mostly because of its thorough focus on constructional patterns and their grounding in social interaction, as well as on the commonalities and differences between the pattern inventories that different individuals have available. While the studies reviewed in the present paper explore first steps towards detecting those patterns, they should be complemented by more qualitatively-oriented follow-up studies that also take the social-interactional aspect into account more thoroughly.

Acknowledgments

We would like to thank the editors of this Yearbook for helpful comments and suggestions. Remaining errors are of course our own.

References

- Ambridge, Ben & Elena Lieven. 2011. *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511975073

- Backus, Ad. 2020. Usage-based approaches. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge handbook of language contact*, 110–126. London & New York: Routledge.
- Bannard, Colin, Elena Lieven & Michael Tomasello. 2009. Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences* 106(41). 17284–17289.
- 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. doi:10.1111/j.1467-9280.2008.02075.x.
- Barron-Hauwaert, Suzanne. 2004. *Language strategies for bilingual families: the one-parent-one-language approach* (Parents' and Teachers' Guides 7). Clevedon: Multilingual Matters.
- Braine, Martin D. S. 1963. The ontogeny of English phrase structure: The first phase. *Language* 39(1). 1–13. doi:10.2307/410757.
- Bruner, Jerome S. 1983. *Child's talk: learning to use language*. Oxford: Oxford University Press.
- Cameron-Faulkner, Thea, Elena Lieven & Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6). 843–873. doi:10.1207/s15516709cog2706_2.
- Clyne, Michael G. 2003. *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.
- Cook, Vivian. 2016. Premises of multi-competence. In Vivian Cook & Li Wei (eds.), *The Cambridge handbook of linguistic multi-competence* (Cambridge Handbooks in Language and Linguistics), 1–25. Cambridge: Cambridge University Press.
- Croft, William. 2001. *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001.
- Croft, William. 2009. Toward a Social Cognitive Linguistics. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New Directions in Cognitive Linguistics*, vol. 24 (Human Cognitive Processing), 395–420. Amsterdam & Philadelphia: John Benjamins.
- Crystal, David. 2003. *English as a global language*. Cambridge & New York: Cambridge University Press. doi:10.1017/CBO9780511486999.
- Dąbrowska, Ewa. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2(3). 219–253. doi:10.1075/lab.2.3.01dab.
- Dąbrowska, Ewa & Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16(3). 437–474. doi:10.1515/cogl.2005.16.3.437.
- Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.
- Gaskins, Dorota, Maria Frick, Elina Palola & Antje Endesfelder Quick. 2021. Towards a usage-based model of early code-switching: Evidence from three language pairs. *Applied Linguistics Review* 12(2). 179–206. doi:10.1515/applirev-2019-0030.
- Goldberg, Adele. 2019. *Explain me this: creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Grosjean, François. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36(1). 3–15. doi:10.1016/0093-934X(89)90048-5.
- Hartmann, Stefan & Antje Endesfelder Quick. 2021. ... und das da go to school. Bilingualer Spracherwerb und Phänomene mehrsprachigen Sprechens. *Der Deutschunterricht* 73(5). 27–36.

- Hartmann, Stefan, Nikolas Koch & Antje Endesfelder Quick. 2021. The traceback method in child language acquisition research: Identifying patterns in early speech. *Language and Cognition* 13(2). 227–253. doi:10.1017/langcog.2021.1.
- Hummel, Katja. 2006. *The traceback handbook*. Unpublished manuscript.
- Ibbotson, Paul, Vsevolod Salnikov & Richard Walker. 2019. A dynamic network analysis of emergent grammar. *First Language* 39(6). 652–680. doi:10.1177/0142723719869562.
- Koch, Nikolas. 2019. *Schemata im Erstspracherwerb: eine Traceback-Studie für das Deutsche* (Linguistik, Impulse & Tendenzen 80). Berlin & Boston: De Gruyter. doi:10.1515/9783110623857.
- Koch, Nikolas, Stefan Hartmann & Antje Endesfelder Quick. 2020. The traceback method and the early construction: theoretical and methodological considerations. *Corpus Linguistics and Linguistic Theory*. doi:10.1515/cllt-2020-0045.
- Kol, Sheli, Bracha Nir & Shuly Wintner. 2014. Computational evaluation of the Traceback Method. *Journal of Child Language* 41(1). 176–199. doi:10.1017/S0305000912000694
- Lakoff, George. 1990. The Invariance Hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1). 39–74. doi:10.1515/cogl.1990.1.1.39.
- Lieven, Elena. 2014. First language development: A usage-based perspective on past and current research. *Journal of Child Language* 41(S1). 48–63. doi:10.1017/S0305000914000282.
- Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30(2). 333–370. doi:10.1017/S0305000903005592.
- Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3). doi:10.1515/COGL.2009.022.
- MacWhinney, Brian. 2015. Emergentism. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 689–706. Berlin & New York: De Gruyter.
- McCauley, Stewart M. & Morten H. Christiansen. 2017. Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science* 9(3). 637–652. doi:10.1111/tops.12258.
- McCauley, Stewart M. & Morten H. Christiansen. 2019. Language learning as language use: A cross-linguistic model of child language development. *Psychological Review* 126(1). 1–51. doi:10.1037/rev0000126.
- 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> (accessed 12 September 2022).
- Otheguy, Ricardo, Ofelia García & Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6(3). 281–307. doi:10.1515/applirev-2015-0014.
- Quick, Antje Endesfelder, Elena Lieven, Ad Backus & Michael Tomasello. 2018a. Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism* 8(3). 393–409. doi:10.1075/lab.17008. qui.
- 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, Elena Lieven, Malinda Carpenter & Michael Tomasello. 2018c. Identifying partially schematic units in the code-mixing of an English and German speaking child. *Linguistic Approaches to Bilingualism* 8(4). 477–501. doi:10.1075/lab.15049.qui.
- Quick, Antje Endesfelder & Stefan Hartmann. 2021. The building blocks of child bilingual code-mixing: A Cross-Corpus Traceback Approach. *Frontiers in Psychology* 12. 2203. doi:10.3389/fpsyg.2021.682838.
- Quick, Antje Endesfelder, Stefan Hartmann, Ad Backus & Elena Lieven. 2021a. Entrenchment and productivity: The role of input in the code-mixing of a German-English bilingual child. *Applied Linguistics Review* 12(2). 225–247. doi:10.1515/applrev-2019-0027.
- Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2021b. Entrenchment effects in code-mixing: Individual differences in German-English bilingual children. *Cognitive Linguistics* 32(2). 319–348. doi:10.1515/cog-2020-0036.
- Quinto-Pozos, David. 2009. Code-switching between sign languages. In Barbara E. Bullock & Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*, 221–238. Cambridge: Cambridge University Press.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press. doi:10.1093/oso/9780198814771.001.0001.
- Stoll, Sabine, Kirsten Abbot-Smith & Elena Lieven. 2009. Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science* 33(1). 75–103. doi:10.1111/j.1551-6709.2008.01004.x
- Taylor, John R. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.
- Tomasello, Michael. 1992. *First verbs. A case study of early grammatical development*. Cambridge: Cambridge University Press.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA & London: Harvard University Press.
- Tomasello, Michael. 2009. The usage-based theory of language acquisition. In Edith Laura Bavin (ed.), *The Cambridge handbook of child language*, 69–87. Cambridge: Cambridge University Press.
- Tomasello, Michael & Elena Lieven. 2008. Children's first language acquisition from a usage-based perspective. In Peter Robinson & Nick J. Ellis (eds.), *Handbook of Cognitive Linguistics and second language acquisition*, 168–196. New York & London: Routledge.
- Tomasello, Michael & Daniel Stahl. 2004. Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* 31(1). 101–121.
- Vihman, Marilyn May. 1999. The transition to grammar in a bilingual child: Positional patterns, model learning, and relational words. *International Journal of Bilingualism* 3(2–3). 267–301. doi:10.1177/13670069990030020801.
- Vogt, Paul & Elena Lieven. 2010. Verifying theories of language acquisition using computer models of language evolution. *Adaptive Behavior* 18(1). 21–35. doi:10.1177/1059712309350970.

