DE GRUYTER MOUTON Corpus Linguistics and Ling. Theory 2022; 18(3): 477–504

Article

Nikolas Koch*, Stefan Hartmann and Antje Endesfelder Quick

The traceback method and the early constructicon: theoretical and methodological considerations

https://doi.org/10.1515/cllt-2020-0045
Received July 20, 2020; accepted November 23, 2020; published online December 14, 2020

Abstract: Usage-based approaches assume that children's early utterances are item-based. This has been demonstrated in a number of studies using the traceback method. In this approach, a small amount of "target utterances" from a child language corpus is "traced back" to earlier utterances. Drawing on a case study of German, this paper provides a critical evaluation of the method from a usage-based perspective. In particular, we check how factors inherent to corpus data as well as methodological choices influence the results of traceback studies. To this end, we present four case studies in which we change thresholds and the composition of the main corpus, use a cross-corpus approach tracing one child's utterances back to another child's corpus, and reverse and randomize the target utterances. Overall, the results show that the method can provide interesting insights—particularly regarding different pathways of language acquisition—but they also show the limitations of the method.

Keywords: construction grammar; language acquisition; traceback; usage-based model

1 Introduction

From the early 2000s on, the traceback method has played a substantial role in corpus-based approaches to child language acquisition. Pioneered by Lieven

***Corresponding author: Nikolas Koch**, Institute for German as a Foreign Language, Ludwig-Maximilians-Universität München, München, Germany, E-mail: koch@daf.lmu.de
Stefan Hartmann, German Department, Heinrich-Heine-Universität Düsseldorf, Dusseldorf, Germany, E-mail: hartmast@hhu.de
Antje Endesfelder Quick, British Studies, University of Leipzig, Leipzig, Germany, E-mail: antje.quick@uni-leipzig.de

et al. (2003) and Dąbrowska and Lieven (2005), the basic idea of the method is that the child's utterances in a small test corpus—usually the last session(s) of recording—are "traced back" to previous utterances in order to demonstrate that a large proportion even of a child's more complex utterances can be accounted for with the help of a finite set of patterns. These patterns can either be exact repetitions of previous material (so-called fixed strings) or partially schematic units with an open slot like [*What's* X], so-called frame-and-slot patterns.

The results of traceback studies have repeatedly shown that the vast majority of utterances can be successfully constructed from fixed strings or frame-and-slot patterns identified in the previous discourse. In recent work, the traceback method has also served as a starting point for more refined quantitative and computational approaches to child language. Building on the basic idea of the traceback procedure, McCauley and Christiansen (2017, 2019) provide computational evidence that children's early utterances rely on "item-based learning through online processing of simple distributional cues" (p.1).

The aim of the present paper is to (re-)evaluate the method from a corpus-linguistic and usage-based point of view (see Kol et al. 2014 for a critical evaluation from a computational-linguistic perspective). We will first give an overview of the method and the results of previous traceback studies in Section 2, before we replicate the study by Koch (2019), which is summarized in Section 3. We implement a number of manipulations to investigate how they influence the traceback results (Section 4):

1) changing the thresholds which establish a sequence as an "entrenched" unit;
2) changing the composition of the main corpus—including or excluding the caregivers' input;
3) manipulating the main corpus by doing cross-corpus studies;
4) using reversed and randomized versions of the test corpora, following Kol et al. (2014).

We argue that all four aspects can lead to a more thorough understanding of the method and its limitations: As will be discussed in more detail in Section 2, the method relies on arbitrary thresholds. It is obvious that these will have an impact on the proportion of successfully traced-back utterances. The extent of this impact is what we would like to address with our first manipulation. The second manipulation touches on theoretical issues that have been discussed multiple times in previous traceback studies: On the one hand, including the caregivers' input may entail the risk that the children are credited with linguistic knowledge that they do not possess (as argued by Ambridge and Lieven 2011)—on the other hand, children's early language use is strongly influenced by the input they receive (see, e.g., Behrens 2006), and as child language corpora only cover a very small proportion of a child's linguistic experience, leaving the input out might make the method overly conservative. The third manipulation allows us to check how similar different children's utterances are and to what extent contextual factors potentially influence the results of traceback studies. And the fourth manipulation checks whether the method overgeneralizes in the sense that it is able to derive patterns that have no basis in the actual data.

Our study thus adds to previous findings in three ways: Firstly, we add new manipulations that have not been applied previously; secondly, we apply these manipulations to data from a different language, namely German; and thirdly, as explained above, we reflect on the results from a usage-based perspective. In doing so, we wish to provide an empirical perspective on the potential and limitations of the method.

2 The traceback method

2.1 General assumptions and their operationalization

The traceback method has been developed to test the hypothesis that even highly creative utterances in early child language can be accounted for on the basis of a fairly limited set of "component units", some of which are lexically fixed (*what's this?*), while others consist of a lexically fixed part and an open slot ([*what's* X?]). While the gist of the traceback method remained the same, it has seen various stages of development over time (Dąbrowska and Lieven 2005; Lieven et al. 2003, 2009; Vogt and Lieven 2010). Due to these developments, the applications of the method differ along various "parameters",[1] some of which will be discussed in the subsequent sections. Here we follow Koch's (2019) operationalization of the method, as this is the study that will be reanalyzed below.

In traceback studies, a child language corpus is split in two parts, a smaller test corpus, comprising the last two sessions of recording, and a main corpus, which consists of the remaining corpus data. The child's multiword utterances in the test corpus are used as the so-called target utterances: They are traced back to the main corpus, i.e., for each target utterance, the closest match is identified in the main corpus. In the simplest case, a component unit is identical to the target utterance. For example, if the target utterance is *what's this?* and the exact same utterance can be found at least twice in the main corpus, the traceback is considered successful, and it is (tentatively) assumed that the child has the fixed chunk *what's this*?

1 We use the terms "parameter" and "parameter settings" in a theory-neutral way here as it seems to be a good metaphor for the methodological choices that the analyst has to make in applying the traceback method.

readily available as an entrenched unit.[2] If no verbatim match can be found, the next step is to check whether it can be derived from potential component units in the main corpus via a set of predefined operations. In the present paper, three operations are used: ADD, SUBSTITUTE, and SUPERIMPOSE (see Table 1).

Consider, for example, a target utterance like *Ich will den Teddy* "I want the teddy". First, the closest match is identified in the main corpus. If the same utterance occurs verbatim in the main corpus, it is considered a fixed string. If the search for a matching fixed string is not successful, the next step is searching for a frame-and-slot pattern. In our example, an utterance in the main corpus could be *Ich will den Kuchen* "I want the cake", which shares a substantial amount of lexical material with the target utterance. Thus, the utterance can be derived with the help of a SUPERIMPOSE operation (see Table 1), which allows us to posit the frame-and-slot pattern [*Ich will den* X]. However, it is also usually assumed that such frame-and-slot patterns are semantically constrained. This is why most traceback studies work with a set of semantic slot categories such as REFERENT, PROCESS, etc.

Table 1: Operations assumed for deriving target utterances from utterances in the main corpus (adopted from Koch 2019: 180).

Operation	Description	Example
SUPERIMPOSE	One component unit is inserted into the open slot of a schema. There is lexical overlap between the schema and the component unit.	Schema: [*Ich will den* REF] "I want the REF" Component unit: *den Teddy* "the teddy" → *Ich will den Teddy* "I want the teddy"
SUBSTITUTE	A component unit is inserted into the open slot of a schema	Schema: [*das ist* ATT] "this is ATT" Component unit: *hoch* "high" → *das ist hoch* "this is high"
ADD	Linear juxtaposition of component units	Component unit 1: *geh weg* "go away" Component unit 2: *Mama* "Mummy" → *geh weg Mama!* "Go away Mummy!"

2 Note, however, that especially Lieven et al. (2003) are very cautious in their interpretation of traceback results when it comes to cognitive plausibility.

Table 2: Types of slots (examples from Vogt and Lieven 2010).

Type of slot	Example utterances		Schema with slot
REFERENT	CHI	More choc choc **on there.**	**REFERENT on there**
	CHI	Bow-'s food **on there.**	
PROCESS	CHI	**I want to** get it.	**I want to PROCESS**
	MOT	And **I want to** talk to you about the park.	
ATTRIBUTE	CHI	Pilchard there **he's** hungry @sc toast.	**He's ATTRIBUTE**
	CHI	**He's** upside+down.	
LOCATION	CHI	**I sit** on Mummy-'s bike.	**I sit LOCATION**
	CHI	**I sit** there.	
DIRECTION	CHI	**Going** under bridge.	**Going DIRECTION**
	CHI	**Going** down.	

Here, the schema would therefore be [*Ich will den* REFERENT]. The semantic slot types assumed in traceback studies vary considerably—in this paper, we follow Koch (2019) in using five slot categories, as shown in Table 2.

2.2 Implementation

In most traceback studies, the method has been applied (semi-)manually. For example, Lieven et al. (2009) and Vogt and Lieven (2010) use an algorithm to automatically identify potential component units in the main corpus. However, frame-and-slot patterns are identified via manual annotation of the data. Kol et al. (2014), by contrast, have used a fully automated traceback procedure that has been adapted and modified by Koch (2019) and will also be used in the present paper.

The traceback algorithm used here is based on the idea that part-of-speech annotations can provide a rough proxy to the semantic slot types. Specifically, the algorithm works as follows:

1. For each target utterance, the algorithm first searches for an exact match in the component units available in the main corpus. A multiword unit counts as a component unit if it is attested at least twice in the main corpus. If an exact match is found, the traceback procedure is considered successful with 0 operations.[3]
2. If no exact match can be found, the algorithm continues looking for partial matches, more specifically: for the largest available partial match. Consider a target utterance like *will lieber das haben* "rather wanna have that":

3 Following previous traceback studies, exact repetitions of material in the preceding five utterances were not traced back (see Koch 2019: 172–173).

The algorithm searches for so-called component units in the main corpus from which this particular utterance can be "constructed". In particular, the algorithm looks for

a. component units that contain a part of the target utterance and a string with a semantically matching POS annotation **before** the match, e.g., *mag lieber das haben* "rather wish to have that", in which case a frame-and-slot pattern [PROCESS *lieber das haben*] could be posited. The traceback is considered successful if the slot filler—in this case *mag*—is also attested at least twice in the main corpus and thus qualifies as a component unit.
b. Component units that contain a part of the target utterance and a string with a semantically matching POS annotation **after** the match, e.g., *will lieber das essen* "rather wanna eat that", in which case a frame-and-slot pattern [*will lieber das* PROCESS] could be posited. The traceback is considered successful if the slot filler—in this case *essen*—is also attested at least twice in the main corpus and thus qualifies as a component unit.
c. Component units that contain a part of the target utterance and a string with a semantically matching POS annotation **in-between** the matching parts, e.g., *will lieber ein Eis haben* "rather wanna have an ice cream", in which case a frame-and-slot pattern [*will lieber* REFERENT *haben*] could be posited. The traceback is considered successful if the slot filler—in this case either the entire NP *ein Eis* or, e.g., the frame-and-slot pattern [*ein* N] and the noun *Eis*—is also attested at least twice in the main corpus and thus qualifies as a component unit.
d. Component units that contain the lexically matching part of the target utterance in-between two strings whose POS annotation matches the POS annotation of the target utterance, which leads to the assumption of a frame-and-slot pattern with two open slots. For instance, *mag lieber das essen* "rather wish to eat that" would lead to the schema [PROCESS *lieber das* PROCESS]. The traceback is considered successful if both slot fillers—in this case *mag* and *essen*—are also attested at least twice in the main corpus and thus qualify as component units.

3. If there is a partial match but the target utterance cannot be fully derived with the help of a frame-and-slot pattern, the algorithm checks whether the remaining lexical material in the target utterance is also available as a component unit in the main corpus. If so, they are combined via an ADD operation: For instance, if the target utterance is *will das haben Mama* "wanna have that Mummy" and the string *will das haben* can be derived with the help of steps 1 and 2, then the algorithm checks if *Mama* is a component unit too (i.e., occurs at least twice in the main corpus). If so, the derivation is considered successful, if not, it is considered a failed derivation. Note, however, that ADD is

only allowed if the units that are juxtaposed can occur in either order (a restriction that will be discussed in more detail in Section 4.4 below). To determine whether or not this applies to a given unit, the algorithm again makes use of the POS annotation.[4]

In cases where multiple derivations are possible, a set of rules applies: First, if component units of several length compete with each other, the candidate with the longest consecutive match is preferred. For example, if [*mag lieber* REFERENT *haben*] and [*mag lieber das* PROCESS] compete with each other, the latter wins out. Secondly, following Vogt and Lieven (2010: 24), the combination strategies SUPERIMPOSE and SUBSTITUTE are given preference over ADD, provided that both variants are possible. Vogt and Lieven (2009: 24) motivate this choice with the key role of frame-and-slot patterns that is assumed in usage-based approaches.

Figure 1 shows a simplified visualization of the algorithm, where a simple three-unit target utterance *ABC* is derived (e.g., *What is this?*). The branches show the most important checks that the algorithm performs. The large letters stand for the closest match in the main corpus, while the smaller letters below represent the

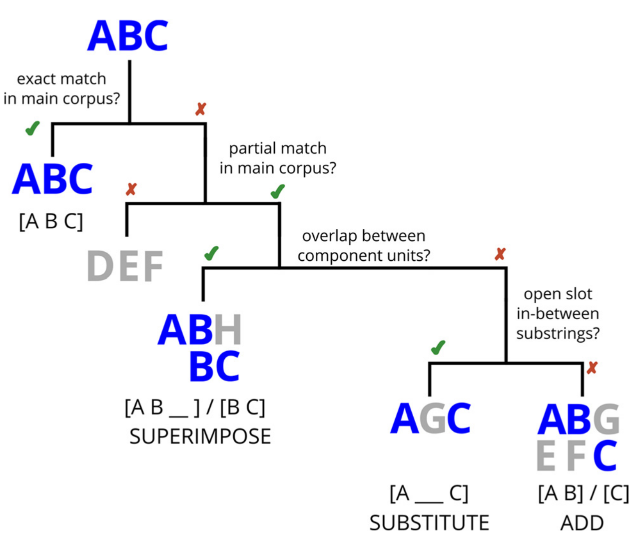

Figure 1: Simpified depiction of the traceback algorithm.

4 More specifically, ADD is only allowed if the unit in question has one of the following POS tags: proper name (n:prop), adverb (adv), conjunction (conj), communicator (co, e.g., *ah, oh, wow*), or onomatopoeia (on).

component units (fixed strings and frame-and-slot patterns) that are posited on the basis of this closest match.[5]

Following Kol et al. (2014), the MOR tags available in the annotated data were used to operationalize the slot categories (see Koch 2019: 186–187). The algorithm used in Koch (2019) as well as the present paper differs from the one used by Kol et al. (2014) in a crucial way: In Kol et al.'s algorithm, a slot can only be filled with one single word. The algorithm used here, by contrast, also allows for phrases as slot fillers. They were identified via combinations of POS tags. The mapping of POS sequences to syntactic categories may not always be unambiguous, but it can be expected to be more reliable in child language (and child-directed speech) than in other varieties of language that tend to be syntactically more complex. Thus, identifying phrases based on POS sequences cannot be expected to significantly increase the number of erroneous matches compared to Kol et al.'s approach of token-based matching.

3 Case study

The study by Koch (2019) was one of the first that applied the traceback method to a language other than English (but see Miorelli 2017 for Italian). Four high-density corpora of German-speaking children aged between 2;0 and 2;6 years were used. Data collection for each corpus took place over a period of seven weeks. Each week, five recordings were made in typical play interactions each lasting 1 h. Table 5 in the Appendix gives an overview of the corpora.[6]

The last two data collection intervals served as the test corpus. The previous recordings were used as the main corpus, taking into account the caretakers' input. As Koch (2019) was mainly interested in the development of syntactic productivity beyond the two-word stage, only utterances longer than two words were used as target utterances in the test corpus. For the sake of comparability, we will follow this approach in the present study.

5 The possibilities displayed in Figure 1 are not exhaustive: For example, the figure should not be taken to imply that frame-and-slot patterns like [AB __] are only possible with SUPERIMPOSE while others with a pattern-initial open slot like [A __ C] are only possible with SUBSTITUTE. Instead, all kinds of frame-and-slot patterns with different levels of complexity can be posited based on both operations.

6 For further details see Koch (2019: 157–170).

Using the method as described in Section 2[7], Koch (2019) showed that between 70% (Merit, N = 340) and 90% (Simon, N = 90) of the four children's multiword utterances in German could be successfully derived. Up to 26% were exact matches, most others could be derived with one single operation or a combination of two operations. The use of three or more operations was only necessary for very few target utterances. The overall results coincide, on the one hand, with the results of the traceback studies on English. On the other hand, the findings are consistent with work by Bannard and Lieven (2009), Bannard and Matthews (2008), Lieven et al. (1992), Pine and Lieven (1993), as well as Tomasello (1992, 2003, 2006), as these studies have also shown that early utterances of children are based on numerous one-to-one repetitions of utterances that they had either produced themselves or found in the input. Table 3 lists the results in more detail. It also includes an analysis of the failed derivation: It is common in traceback studies to distinguish between derivations that fail for purely lexical reasons (i.e., a lexical item does not occur in the main corpus) and tracebacks that fail for syntactic reasons.

Note, however, that the results for Merit differ quite considerably from those for the other children. As she seems to be the most advanced learner generally, this raises the question of whether the traceback success is inversely correlated to the grammatical complexity of a child's utterances. This is why Koch (2019: 210–211) also takes a closer look at the mean length of utterance (MLU) of the individual children (also given in Table 3, calculated from the child's utterances in the test corpus). MLU is often used as a rough measure of grammatical complexity, even though it has its limitations and cannot distinguish between different types of complexity (Clark 2009: 212). Following Behrens (2006: 11; 2016) as well as the previous traceback studies (see Dabrowska and Lieven 2005: 445; Lieven et al. 2009: 485; Vogt and Lieven 2010: 25–26), a word-based MLU measure was used. Indeed, the traceback success tends to be lower for children with a higher MLU, while the proportion of exact matches and successful tracebacks requiring only one single operation tends to be higher for children with lower MLU. An interesting exception is Marieke: Although her MLU value (2.38) was higher than that of Leo (1.62) and Simon (1.82), there were more direct matches (26% vs. 17% for Leo and 10% for Simon). One aspect that seems to play an important role here are qualitative differences between the test corpora. In the case of Marieke, 42% (N = 74) out of 175 target utterances were question constructions. In Simon's case, however, the

7 The results reported here differ slightly from those in the study of Koch (2019): There, an additional manual step was taken to analyze the data (for details see Koch 2019: 189–190). As replicating this step for all the studies presented in Section 4 would have entailed disproportionate effort, we report the results of a fully automated analysis here.

Table 3: Overview of successful tracebacks (Table adopted from Koch 2019: 197).

		Leo MLU: 1.62		**Marieke MLU: 2.38**		**Merit MLU: 4.85**		**Simon MLU: 1.82**	
		Number	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**
Target utterances		113	100	175	100	486	100	100	100
Successful tracebacks		90	80%	141	81%	340	70%	90	90%
Number of operations	Exact Match	15	17%	37	26%	26	7.5%	9	10%
	1	51	57%	73	52%	139	41%	62	69%
	2	23	25%	28	20%	129	38%	17	19%
	3	1	1%	3	2%	39	11.5%	1	1%
	4	0	0%	0	0%	7	2%	1	1%
	5	0	0%	0	0%	0	0%	0	0%
	> 5	0	0%	0	0%	0	0%	0	0%
Types of operations	SUPERIMPOSE	32	28.3%	43	24.6%	133	27.4%	24	24%
	SUBSTITUTE	62	54.9%	84	48.0%	400	82.3%	79	79%
	ADD	6	5.3%	12	6.86%	14	2.88%	4	4%
Fails	Lexical	11	9.73%	22	12.6%	94	19.3%	5	5%
	Syntactic	12	10.6%	11	6.29%	51	10.5%	4	4%

proportion of question constructions was only 14% (N = 14), for Merit, it was 15% (N = 74), and in Leo's case, no question constructions occurred at all. Question constructions are clearly more formulaic structures that occur with high frequency in specific contexts.[8] For this reason, we assume that the higher proportion of direct matches in the Marieke corpus can be attributed to the increased occurrence of question constructions in her test corpus.

This indicates that the results of traceback studies are highly dependent on the composition of the test corpus. At the same time, the composition of the main corpus plays a crucial role as well. For instance, Lieven et al. (2009: 492), who compared traceback results of four children at the same MLU stage (rather than the same age), attribute the individual differences that they found to differences in the degree to which children rely on what has already been said or perceived in the input. Thus, the question of whether or not the input is included in the main corpus (which some traceback studies do, while others don't) might have a significant impact on traceback results.

These considerations are also connected to an important limitation of the traceback method. As mentioned above, its theoretical preassumption is that *all* of children's utterances are somehow based on previously encountered material. The primary goal of the traceback approach is to substantiate this hypothesis by showing that indeed most target utterances can be accounted for in a systematic and principled way. But it is self-evident that not all target utterances can be traced back successfully (because we are only dealing with samples) and that the proportion of successfully traced-back utterances depends on various factors like sample size and sample composition (which includes contextual factors that influence which words and constructions are attested in the corpus). In Section 4, we will therefore also take a closer look at how the composition of the corpus affects the results of traceback studies.

However, not only the composition of the corpus plays a role, but also the way in which the traceback procedure is operationalized. In Koch (2019: 244–265), an analysis of the failed derivations showed that more than half of them (62 out of 103) could be attributed to the relatively restricted application of the ADD operation (see point 3 in the description of the algorithm in Section 2.2). A theoretical motivation for this restriction couched in the framework of Cognitive Grammar is provided by Dąbrowska (2004: 215).[9] On this view, the relationship between two components that are combined via juxtaposition is "paratactic rather than truly

8 On the role of frame-and-slot patterns in question constructions, also see Rowland (2007), who shows that many errors in children's question constructions can be attributed to problems filling the variable slots of the frame (e.g., *Does she has ears?, Why did they ran away?*).

9 Thanks to an anonymous reviewer for pointing this out to us.

syntactic: the fact that they are placed next to each other signals that their meanings are to be integrated, but the grammar does not spell out how this is to be done, so it must be inferred by the listener." (Dąbrowska 2004: 215)

Note that these restrictions on ADD yield even more conservative results if the method is applied using a computational algorithm: When working manually, one could rely on native-speaker intuitions in judging whether a component unit could occur in either position. When using an automated algorithm, specific criteria have to be defined for allowing or disallowing an operation.

Kol et al. (2014), however, question these restrictions, arguing that "it is unclear how these restrictions are determined, or even how the child could know them". While this is indeed an important conceptual question, their methodological choice to drop the restrictions in their own study entails that the method in the way they apply it will almost certainly lead to more successful derivations than the more constrained version used in previous studies. In Section 4.4 below, we will adopt their approach, but we retain the more conservative operationalization of the ADD operation from previous traceback studies.[10]

Summing up, the case study presented in this section highlights some strengths of the traceback method while also pointing to potential weaknesses that will be explored in more detail in Section 4. Perhaps most importantly, it lends further support to the hypothesis that children are "creative only in highly constraining ways" (Tomasello 2007a: 839), constructing utterances on the basis of already acquired structures using a "cut-and-paste" strategy, as Tomasello (2007b: 1007) calls it. The open questions that the study raises with regard to the traceback method boil down to the overarching question of how different parameter settings—e.g., the choice of main and test corpus and the composition of the corpora—influence the outcome of traceback studies.

4 (Re-)Evaluating the traceback method: The impact of parameter settings

In this section, we explore to what extent different parameter settings influence the outcome of the traceback method. To this end, we use the data introduced in Section 3 and manipulate individual parameters: We change the thresholds for identifying component units (Section 4.1) and the composition of the main corpus

10 Note that we make this choice for reasons of comparability: Kol et al.'s (2014) question of whether the restrictions on ADD are psychologically plausible still stands, but answering it is beyond the scope of the present paper.

(Section 4.2); we swap the main corpora across the four children (Section 4.3); and we reverse and randomize the target utterances (Section 4.4).

4.1 Changing thresholds

In our first manipulation study, we changed the minimum frequency that is required for a fixed string or frame-and-slot pattern in the main corpus to be accepted as a component unit. In their traceback study investigating the acquisition of English question constructions, Dąbrowska and Lieven (2005: 447) have established a frequency threshold of two occurrences in the main corpus for identifying component units, "assuming that children have stored all units that occur with a frequency of 2 or more in the corpus". They further discuss whether this threshold is cognitively realistic, arguing that the corpora they investigate represent only a small fraction of what the children hear and produce over the recording period, which is why the actual number of occurrences of the respective chunks and schemas can be expected to be much higher in the child's actual linguistic experience (Dąbrowska and Lieven 2005: 456–457). In other words, their goal in applying the traceback method is to account for as many of the target utterances as possible with the help of patterns that can be considered (more or less strongly) entrenched. Raising the frequency threshold entails that the method is operationalized in a more conservative way.

Dąbrowska (2014: 640–641) has already shown that changing the frequency threshold inevitably leads to an increase in failed derivations, but "the increase in fails is very gentle. This is because […] most of the component units used in the derivation are quite frequent." (Also see Dąbrowska and Lieven 2005: 457 for a similar reanalysis of their data.) Our study substantiates this result, even though German has a significantly higher degree of variation in inflectional morphology compared to English. Thus, it could have been expected that the manipulation of the frequency criterion would yield more notable differences for German.

Figure 2 shows the percentage of successful tracebacks for each threshold manipulation. In all cases, the caregivers' input was taken into account in the main corpus. The dots and lines show the cumulative traceback success, i.e., the position of each datapoint on the *y* axis indicates the overall cumulative percentage of target utterances that could be successfully traced back with the number of operations indicated on the *x* axis. For example, using the threshold of 1 for both fixed strings and schemas (leftmost in Figure 2), 18 of Simon's 100 target utterances

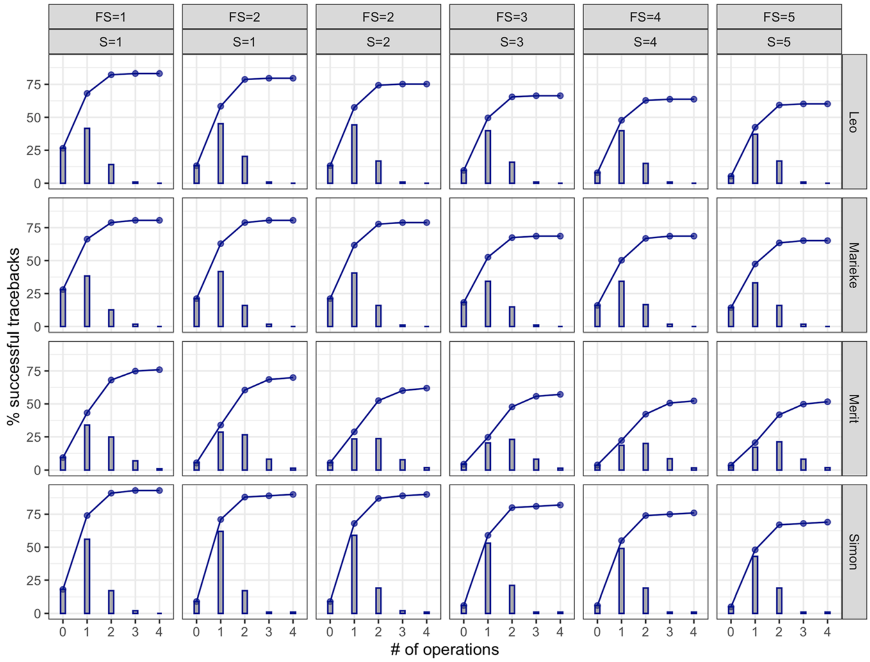

Figure 2: Results of the traceback studies with varying thresholds. The thresholds were set independently for fixed strings and frame-and-slot patterns (schemas): "FS" stands for "fixed string" and "S" for "schema", i.e., *FS=2; S=1* means that the threshold for assuming a component unit is 2 for fixed strings and 1 for lexically specific schemas. The dots and lines show the cumulative traceback success, the bars show the proportion of target utterances successfully derived with the respective number of operations.

yielded exact matches in the main corpus (0 operations). About 56 additional utterances could be traced back with one operation. This is shown by the bars in Figure 2, which indicate the proportion of traceback success for each number of operations individually. About 17 utterances could be derived with two operations (summing up to 91), and two with three operations. Overall, then, 93% of Simon's 100 target utterances could be successfully derived.

As could be expected, the overall traceback success gets lower the higher the threshold is. However, the decline is slower than one might perhaps expect. For obvious reasons, the percentage of direct matches decreases as the thresholds grow bigger, and the proportion of successful derivations that require two or more operations increases. Interestingly, the values are quite consistent in the case of Marieke. As mentioned in Section 3, Marieke's test corpus differs from those of the others in that it has a large number of question constructions, which show lower syntactic and lexical variability.

4.2 Changing the composition of the main corpus—including or excluding caregivers' input

Traceback studies differ in whether or not they take the caregivers' input into account. On the one hand, it could be argued that leaving the caregivers' input out makes the method more cautious and, thus, more reliable. For instance, Ambridge and Lieven (2011: 218) consider the inclusion of input in earlier studies as a "drawback", arguing that the child is "almost certainly credited with schemas that she showed no evidence of possessing." In another study, however, Vogt and Lieven (2010: 23) decided to include the caretaker input as they see this procedure as "closer in spirit" to the theoretical model that informs their approach (as well as the computational model that they use to test it): Usage-based theory assumes that children's learning of constructions is closely related to the input they receive.

In order to check to what extent including or excluding the input influences the traceback results obtained on the basis of the data discussed here, we performed the same traceback study for each of the four corpora twice—first on the basis of a main corpus that includes the input, then using a main corpus that only contains the child's own utterances.

As Figure 3 shows, the traceback results are consistently worse across all corpora if the input is omitted. We fit binomial mixed-effects regression models to the data to check whether the influence is statistically significant.[11] The outcome (success vs. failure) for each target utterance was used as the response variable, the main corpus (with vs. without input) was used as the only predictor variable, and the target utterance was included as a random effect. The resulting coefficient tables for each of the four datasets are shown in Table 4. Across all corpora, the model predicts significantly better results for the main corpus that includes the input.

But even though including the input increases the traceback success across the board, both the model and Figure 4 show that there are individual differences between the children. For instance, the differences between using the main corpus with or without input are most pronounced in the case of Simon. This can potentially be explained by the fact that the number of child utterances in his main corpus is lower than in Merit's case and much lower than the number of child utterances in Marieke's and Leo's main corpora, both regarding the number of utterances and especially regarding the word count (Simon: 14,350 words, Merit: 43,337, Marieke: 41,953, Leo: 30,625; see Table 5 in the Appendix).

11 Thanks to an anonymous reviewer for suggesting this analysis.

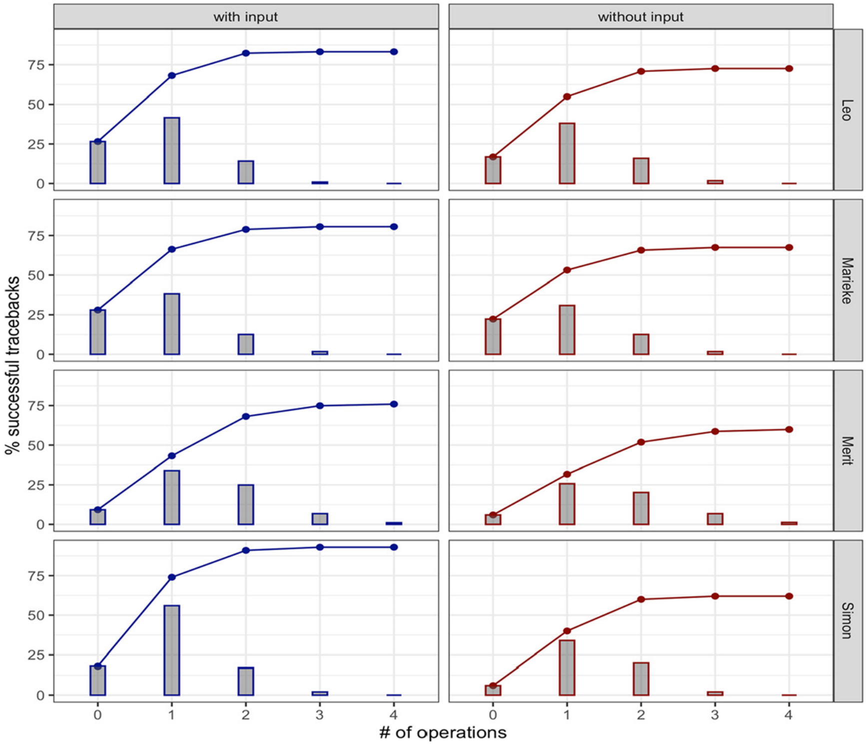

Figure 3: Traceback results for the four corpora under investigation, manipulating the composition of the main corpus (with vs. without input). The dots and lines show the cumulative traceback success, the bars show the proportion of target utterances successfully derived with the respective number of operations.

Table 4: Coefficients of the binomial logistic regression model.

	Estimate	Std. Error	*z* value	Pr(>|*z*|)
Leo				
(Intercept)	10.64	1.32	8.08	6.5E-16[a]
input: with	11.46	1.53	7.49	6.9E-14[a]
Merit				
(Intercept)	1	0.27	3.66	0.00026[a]
input: with	1.23	0.22	5.51	3.5E-08[a]
Marieke				
(Intercept)	9.74	0.98	9.93	2.9E-23[a]
input: with	9.47	1.14	8.3	1.1E-16[a]
Simon				
(Intercept)	0.25	0.31	0.81	0.42
input: with	3	0.78	3.86	0.00011[a]

[a]<0.001.

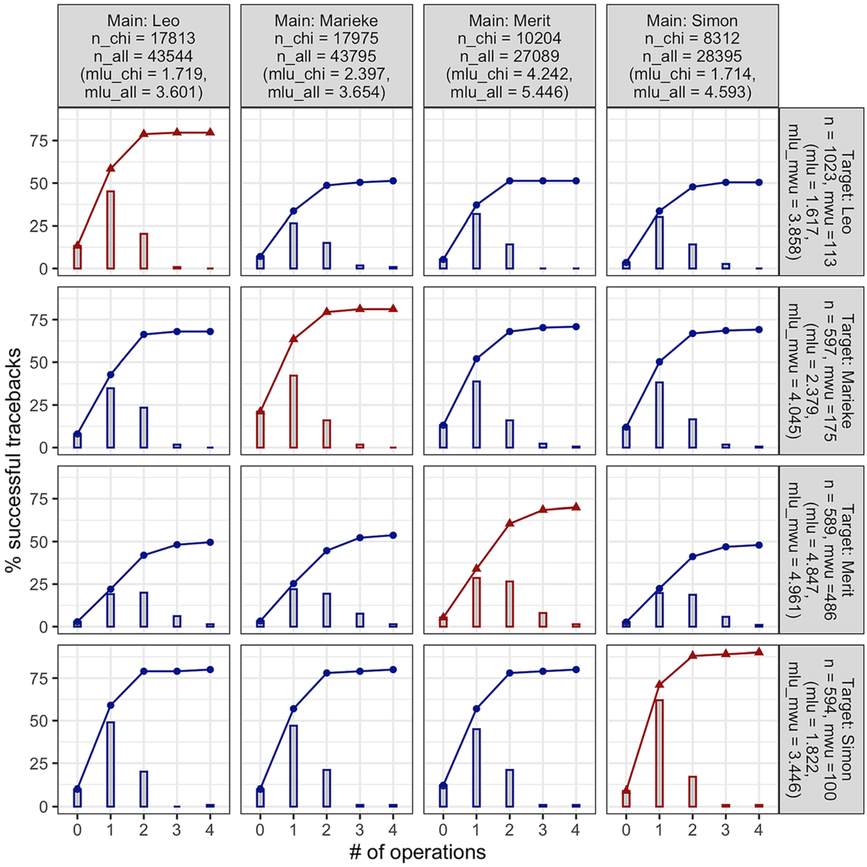

Figure 4: Results of the cross-corpus traceback study. The size of the main and test corpora are given in the headers. “n” stands for the total number of utterances, “mwu” for the number of multiword utterances. The proportion of successful tracebacks is calculated relative to the number of multiword units (defined as utterances with 3 or more words, see main text) in the test corpus, as only multiword target utterances enter the analysis. The dots and lines show the cumulative traceback success, the bars show the proportion of target utterances successfully derived with the respective number of operations.

Returning to the question of whether the input should be included or not, the answer strongly depends on the research question: If we want to map out a child’s grammar, i.e., the inventory of constructions they have available as firmly entrenched chunks or schemas, it makes sense to limit the scope of the investigation to the child’s utterances. In this way, the traceback method has been applied in a number of studies (e.g., Lieven et al. 2009). If, however, we just want to show that children’s utterances can be fully explained with the help of "building blocks" that they have encountered before, then it makes perfect sense to include the input.

But even if the application of the method aims at carving out a "grammar" and modeling the early "construction", there are some arguments why the input should be included: A number of studies have shown that children's constructions are strongly influenced by the input they receive (see, e.g., Behrens 2006; Cameron-Faulkner et al. 2003). We can therefore expect that the input documented in the corpora can be informative about the child's linguistic knowledge to a certain extent. As such, it can be argued that leaving out the input data would be overly restrictive, especially given that the recorded data is only a very small sample of what the child says and hears: Lieven et al. (2009: 493), who work with very dense corpora (using almost the same sampling schedule that was adopted for the data analyzed in the present study), estimate that their corpora capture about 7–10 percent of what the children hear and say. Thus, it could be argued that the input can "stand in" for the remaining 90 to 99 percent to some extent.

Still, it is an important question to what extent the composition of the main corpus influences the outcome of traceback studies. In Section 4.3, we explore this question in more detail by swapping the main corpora between the different children.

4.3 Cross-corpus traceback

In traceback studies, the target utterances of one child are typically traced back to the same child's earlier utterances. For the present study, however, we swap the main corpora across the four children, tracing the utterances in each test corpus back to all utterances in the main corpora established for the child him- or herself as well as the three other main corpora. This can help us answer three different, but interrelated, questions[12]:

a) What happens when we trace back one child's target utterances using other children's main corpora?
b) To what extent do the characteristics of the main corpus influence the traceback results?
c) To what extent do the characteristics of the test corpus influence the traceback results?

Figure 4 provides an overview of the results, which we will now discuss in more detail, focusing on the implications for the three questions raised above.

12 We would like to thank the anonymous reviewer who has suggested this phrasing of a)–c) to us.

First, what happens if we swap the main corpora? For all children, the traceback success is higher within each child's dataset, even if one or more of the other main corpora are considerably bigger. Using mixed-effects logistic regression models (whose results are reported in Appendix C), it can be shown that for all four test corpora, the traceback success decreases significantly when one of the other children's main corpora is used.[13] Thus, the fact that the main corpus comes from the same dataset seems to be more important for traceback success than other factors such as the size of the main or test corpus. A regression model using the pooled data from all cross-corpus studies, also reported in Appendix C, lends further support to this hypothesis: The type of traceback (within-corpus vs. cross-corpus) has the strongest effect on traceback success.

Figure 4 shows the results; the sizes of the test and main corpora are given there as well. *n* indicates the total number of utterances, *mwu* the number of multiword utterances (MWUs). Merit's test corpus with 486 MWUs is much larger than the test corpora of the other three children. As for the main corpora, those containing the utterances of Marieke and Leo as well as their respective caregivers are much larger than the main corpora for Merit and Simon.

The results in Figure 4 show that the size of the main or test corpus may have an impact on traceback success: Simon has the smallest test corpus, and his tracebacks are consistently highly successful; the tracebacks for Merit's target utterances, who has the largest test corpus, are consistently quite low. However, this cannot be the single determining factor: Note, for instance, that for the cross-corpus tracebacks, the percentage of successful tracebacks is similar for Leo's and Merit's target utterances even though Merit's test corpus is more than four times larger. And although Marieke's test corpus is larger than the test corpora of Simon and Leo, the percentage of successful tracebacks is higher than that of Leo and similar to that of Simon. Here it can be instructive to take a look at the different kinds of fails, shown in Figure 5: In the cross-corpus tracebacks, Leo shows a high proportion of lexical fails. A reason for this might be that his vocabulary is very specific and already highly differentiated in the field of trains—the most frequent content words in his target utterances are *fährt* "goes/drives" (36) and *U-Bahn* "subway" (24).

A closer look at the MLU values also yields conflicting results. Although Marieke's target utterances have a higher MLU than those of Leo and Simon, their traceback success hardly differs from that of Simon, while the traceback success is lower in Leo's case. These results suggest that a variety of other factors strongly influence the traceback results. Some candidates include idiolectal characteristics

13 Thanks to an anonymous reviewer for suggesting this analysis.

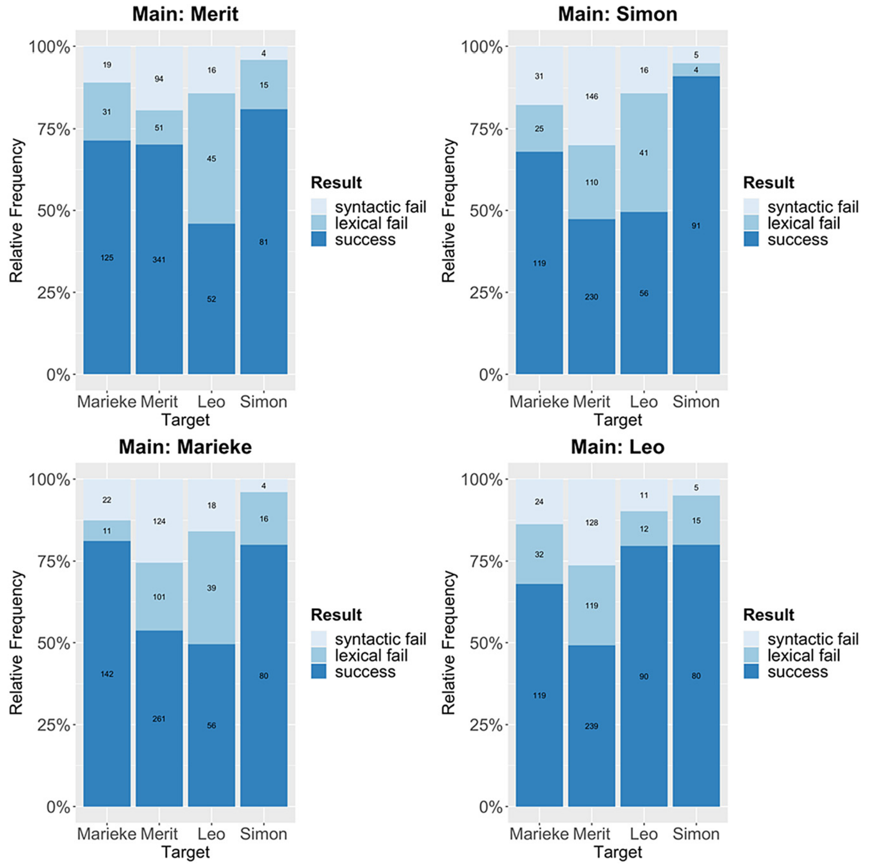

Figure 5: Lexical and syntactic fails in the cross-corpus studies. On the *x* axis, the target children are shown in ascending order of test corpus size. The panels are arranged in ascending order of main corpus size (Merit > Simon > Marieke > Leo).

shared by family members or contextual factors such as recurring topics of conversation as in Leo's case mentioned above.

Overall, however, the number of successful tracebacks of the same test corpus is very similar regardless of the different main corpora. As Figure 4 shows, this is true for all children, no matter how large the main or test corpora are. Furthermore, a large amount of target utterances could be traced back to the other children's main corpora. At first, this might be surprising but we should not forget that the children in our study are fairly young, live similar lives, have a similar background (high socio-economic status) and probably experience similar discourse situations. Although the usage-based account predicts individual differences, we cannot be too different in order to understand each other. As such, it can be assumed that a shared inventory of constructions is acquired, especially in the domain of high-frequency items (also see Street and Dąbrowska 2010). Individual differences are hypothesized to play a more major role for low-frequency types, and these could be the examples where our cross-corpus traceback was less successful than typical within-corpus tracebacks. Also, individual differences are particularly evident in the lexicon, which in turn is strongly influenced by the situational context. Although all recordings were made in play interactions with similar discourse conditions, the importance of the concrete vocabulary used in the recordings for the result of the traceback procedure is evident.

Our results further suggest that the children's language differs in the degree to which they use highly formulaic constructions ("prefabricated" units or prefabs, as e.g., Bybee 2010: 28 calls them). This idea is supported by the differences in traceback success between the similarly sized test corpora of Leo, Marieke, and Simon. As mentioned in Section 3, many of Marieke's utterances are highly formulaic question constructions, which are characterized by a high degree of formulaicity. These structures also occur in the input (of all children) with high frequency. Also note that the traceback success is very high in the case of Simon (with the smallest test corpus), regardless of whose main corpus is used.

As mentioned above, it should be taken into account that the children investigated here share a similar background in terms of various factors including the relatively high socioeconomic status of their families. This might contribute to the observation that their "individual grammars" are similar enough to yield consistently high traceback results even across corpora. The fact that the children investigated in most language acquisition studies come from fairly similar backgrounds is a general problem of the field, and it is an interesting open question whether similar results would be obtained if the children came from more heterogeneous backgrounds.

4.4 Reverse and randomized target utterances

Following up on Kol et al. (2014), we also reversed and randomized the word order in the target utterances. The aim of this procedure was to investigate whether the traceback method can also be used to successfully construct supposedly ungrammatical utterances, which would—according to Kol et al. (2014)—indicate an overgeneralization problem of the method. For example, the utterance *die fährt auf Schienen* "It runs on rails" would become *Schienen auf fährt die* in the reversed dataset and *auf fährt die Schienen* in the randomized dataset.

On the basis of their English dataset, Kol et al. (2014) have shown that the percentage of successful tracebacks remains almost unchanged for the reverse and randomized utterances as compared to the unchanged ones. Kol et al. (2014) concluded that the traceback procedure has a general methodological problem. They attribute this to the overgeneralization potential of the method, which is caused by the ADD operation. However, as mentioned by Koch (2019: 261–262), the construction processes of the manipulated target utterances differ from those of the original ones. For example, in five out of six corpora with modified word order, more than twice as many ADD operations were needed to construct the target utterances. In addition, a number of multiple ADD operations were necessary since there were no fixed word sequences in the main corpus that could be used. In addition, the number of operations needed to derive target utterances increased across all corpora (see Kol et al. 2014: 193). Nevertheless, Kol et al. (2014) were able to show that supposedly ungrammatical utterances could be constructed using the traceback method.

However, there are at least two potential problems with Kol et al.'s study. The first is a methodological one: As already mentioned in Section 2.2, Kol et al. (2014) deviate from previous applications of the traceback method in their operationalization of the ADD operation. In the original traceback studies, the use of the ADD operation is only allowed when the components can occur in either order. Kol et al. (2014) do not seem to constrain the application of ADD in the same way. As pointed out above, the theoretical motivation for this choice is convincing, but for the sake of comparability, we have adopted the more conservative operationalization of the ADD operation for the present study.

The second potential drawback relates to both methodological and theoretical aspects: Kol et al.'s criticism of the method is based on the assumption that mixing up or randomizing the word order would result in ungrammatical target utterances. The successful derivation of these utterances then leads them to the conclusion that the method is overgeneralizing. However, in many cases, randomizing and reversing the target utterances does not necessarily lead to ungrammatical structures and/or structures that are not part of the child's linguistic knowledge.[14] This is especially true for German, which has a less fixed word order than English. In our data, randomization sometimes leads to results that are as

14 Note that "grammaticality"—which is a problematic concept from a usage-based perspective anyway—provides only a rough proxy to the question of whether a structure is actually part of a child's "individual grammar", which is, essentially, what the traceback method is all about. For the sake of the argument, we will however follow Kol et al.'s approach by judging the grammaticality of utterances from the perspective of a codified standard grammar in the remainder of this section.

grammatical/acceptable as the original, as in (2) a and b, or even more acceptable as (2) c. As our randomization algorithm allowed each word order, including the original one, to be sampled, some target utterances also remained unchanged, as in (2) d.

(2) a. *Ich hole das* "I'll get that" (original)—*Das hole ich* (randomized)
  b. *Da liegt das* "It's lying there" (original)—*Das liegt da* (randomized)
  c. *Ja ich mache* "Yes I'll do" (original)—*Mache ich ja* (randomized)
  d. *Cremé gar nicht da* "creme not there at all" (original)—*Cremé gar nicht da* (randomized)
  (all examples from Simon)

The observation that many of the utterances remain or become grammatically correct also applies to the reverse utterances, where in (3) a–c, both variants are equally grammatical, and in d, reversing the word order turns an ungrammatical utterance into a grammatical one.

(3) a. *Ich hole das* "I get/fetch this" (original)—*Das hole ich* (reversed)
  b. *So bleibt es* "So it stays" (original)—*Es bleibt so* (reversed)
  c. *Fina geht da* "Fina goes there" (original)—*Da geht Fina* (reversed)
  d. *Ja ich mache* "Yes I do" (original)—*Mache ich ja* (reversed)
  (all examples from Simon)

Let us now discuss the specific results in more detail. Using reversed and randomized utterances leads to a pronounced decrease in traceback success: In the case of the Simon, for example, only 66% of the target utterances with random word sequences could be derived successfully. This is 24% less than with normal word order. In the case of the randomized test corpus, 69% could be derived successfully. The same tendency holds true for the other three corpora. Similar to Kol et al.'s (2014) study on English data, the number of operations that are needed to construct the target utterances increases for the reversed and randomized data. As Figure 6 shows, there are less verbatim repetitions (0 operations) and less derivations that require only 1 operation but many more that require two operations. Table 6 in the Appendix gives a more detailed overview of the results, listing the operations required for the successful tracebacks as well the types of fails for the ones that were not successful. Expectably, reversing or randomizing the target utterances yields an increase in syntactic fails. Also, the proportion of SUPERIMPOSE operations decreases as this operation is word-order sensitive; however, the

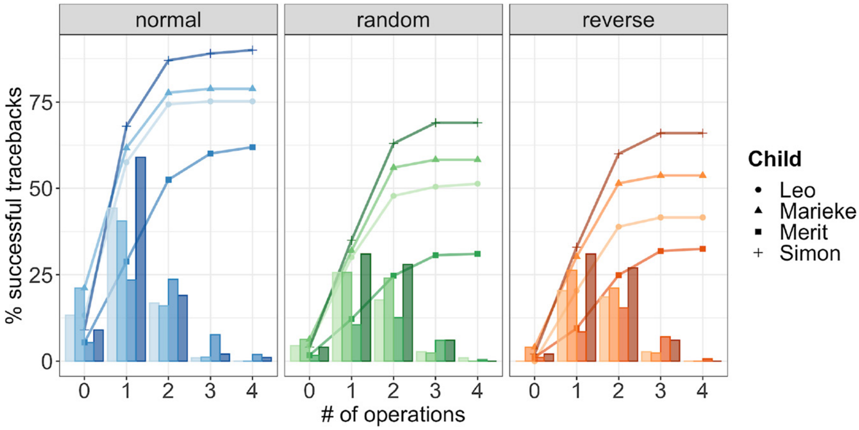

Figure 6: Results of the analysis using target utterances with randomized and reversed word order (threshold of 2 occurrences in the main corpus both for fixed strings and schemas).

restrictive operationalization of ADD prevents the increase in the proportion of ADD operations that could be observed in Kol et al.'s (2014) study.

In sum, a qualitative as well as a quantitative difference can be observed compared to the derivation of target structures with normal word order: On the one hand, fewer target utterances could be constructed (quantitative difference), and on the other hand, the successful construction of target utterances required more steps (qualitative difference). It should be added, however, that the differences between our findings and those by Kol et al. (2014) might not be due to the methodological differences alone. Instead, typological differences between the languages under investigation may play a role as well. As mentioned above, German has a more flexible word order than English. The question of how such typological differences affect traceback results would merit a study in its own right. Note that, for instance, Miorelli (2017) significantly alters the traceback procedure to take the more flexible word order of Italian into account by introducing a new category of component units, "semantic packets" in which the constituents can occur in varying orders. Following Kol et al.'s line of reasoning, however, it could be argued that this category exacerbates any overgeneralization problems that the method may have.

5 Conclusion

In this paper, we have provided a (re)evaluation of the traceback method that has been used in a number of highly influential papers on first language acquisition.

In sum, all manipulation studies presented in this paper lend further support to the main gist of the original traceback studies: In all cases, a large proportion of the target utterances could be traced back successfully, which suggests that children's utterances—even seemingly creative ones—are highly formulaic. But we have also argued that the exact degree to which the traceback procedure is successful strongly depends on contextual factors—including, for example, the situational contexts in which the recordings were made. Our manipulations support this idea: For example, the overall traceback success tends to be lower when another child's main corpus is used. This is of course also closely connected to the "feedback loop" between individual and population that has received increased interest in usage-based theory over the last few years (see, e.g., Petré and Anthonissen 2020; Schmid 2020). Children—and adults—show significant individual differences in language acquisition and use, but at the same time, it is self-evident that a large part of the individual's repertoire must be shared by other speakers for a language to work properly. This balance between idiolectal and shared aspects is arguably also reflected in the cross-corpus traceback results: The fact that the traceback success remained fairly high even when other children's data were used as main corpus indicates that much of the linguistic repertoire is shared between individual speakers. On the other hand, there are interesting differences between the traceback results obtained for each of the four children. These can be partly interpreted in terms of individual differences. Both for the within-corpus and the cross-corpus studies, some of the differences in traceback success can almost certainly be attributed to differences in the children's linguistic development—for example, the traceback success tends to be lower for more advanced learners. Bannard et al. (2009), who used a "trace-forward" approach that uses child corpus data to predict unseen (withheld) utterances in a data-driven way, made a similar observation: Models using lexically specific grammars perform better at predicting earlier data, while models that allow for a higher degree of generalization by using part-of-speech annotations and semantic slots are needed to predict the data of more advanced learners. This is in line with the usage-based idea that children's early language use is first characterized by lexically specific constructions that are acquired by direct imitation before they start using constructions ever more productively and creatively (see, e.g., Bannard et al. 2009: 17289). These developmental processes are of course not directly reflected in traceback studies, as the target utterances represent the child's state of linguistic knowledge at one particular point in time. They only become apparent if we compare multiple traceback results, e.g., for the same child at different stages of their development (as in Lieven et al. 2009) or for different children who may have roughly the same age but still differ in their linguistic development, as in the present study.

This also reminds us that the interpretability of traceback results is limited. Individual traceback studies cannot tell us much about children's linguistic development—only a comparison between different tracebacks can reveal insights beyond the mere proportion of successful tracebacks (which in turn partly depends on the composition of the corpus).

The results of our cross-corpus traceback study can also be interpreted in terms of individual differences as it yields quite different results, depending on whose data are traced back to whose main corpus. While these differences might partly be due to chance factors, they do point to actual differences, as a closer qualitative look at the data reveals. For example, the surprising observation that Marieke's main corpus consistently yields high traceback success can be explained by the fact that she uses more (formulaic) question constructions than the other children.

Thus, the method can prove quite useful even for purposes that go beyond its initial objective to show *that* (but not necessarily *how*) children's language is highly formulaic. As its application requires that patterns are detected in the data, it entails the further advantage that these concrete patterns provide clues to entrenched chunks and lexically specific schemas. However, we have to be wary of any overinterpretation. As the traceback method works with arbitrary frequency thresholds, the detected patterns probably differ significantly in the degree to which they are entrenched. Thus, Kol et al.'s (2014) criticism that the method does not take frequency effects into account is warranted, and if our goal is to draw conclusions about entrenchment, other methods should be chosen, or the traceback method should at least be accompanied by other empirical methods and/or in-depth qualitative analyses of the data. As for the future development of the method, both Bannard et al.'s (2009) trace-forward approach and McCauley and Christiansen's (2017, 2019) "chunk-based learner" represent highly promising first steps toward a more fine-grained model of how linguistic utterances are "constructed" in early child language.

References

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

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. 241–248.

Bannard, Colin & Elena Lieven. 2009. Repetition and reuse in child language learning. In Roberta Corrigan, Edith A. Moravcsik, Ouali Hamid & Kathleen Wheatley (eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, functional explanations*, 297–232. Amsterdam: John Benjamins.

Bannard, Colin, Elena Lieven & Michael Tomasello. 2009. Modeling children's early grammatical knowledge. Proceedings of the National Academy of Sciences 106(41). 17284–17289.

Behrens, Heike. 2006. The input-output relationship in first language acquisition. *Language and Cognitive Processes* 21. 2–24.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Cameron-Faulkner, Thea, Elena Lieven & Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6). 843–873.

Clark, Eve V. 2009. *First language acquisition*, 2nd edn. Cambridge: Cambridge University Press.

Dąbrowska, Ewa. 2004. *Language, mind and brain: Some psychological and neurological constraints on theories of grammar*. Edinburgh: Edinburgh University Press.

Dąbrowska, Ewa & Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16(3). 437–474.

Koch, Nikolas. 2019. *Schemata im Erstspracherwerb: Eine Traceback-Studie für das Deutsche*. Berlin, Boston: De Gruyter.

Kol, Sheli, Bracha Nir & Shuly Wintner. 2014. Computational evaluation of the Traceback method. *Journal of Child Language* 41(1). 176–199.

Lieven, Elena, Julian M. Pine & Helen D. Barnes. 1992. Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language* 19(2). 287–310.

Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30. 333–370.

Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20. 481–507.

McCauley, Stewart M. & Morten H. Christiansen. 2017. Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science* 9(3). 637–652.

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.

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.

Petré, Peter & Anthonissen Lynn. 2020. Individuality in complex systems: A constructionist approach. *Cognitive Linguistics* 31(2). 185–212.

Pine, Julian M. & Lieven Elena. 1993. Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language* 20. 551–571.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system. Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.

Street, James A. & Dąbrowska Ewa. 2010. More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120(8). 2080–2094.

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: Harvard University Press.

Tomasello, Michael. 2006. Acquiring linguistic constructions. In William Damon, Richard M. Lerner, Deanna Kuhn & Robert S. Siegler (eds.), *Handbook of child psychology*, vol. 2: *Cognition, perception, and language*, 255–298. New York: Wiley.

Tomasello, Michael. 2007a. A usage-based approach to child language acquisition. In Benjamin Bergen, Vyvyan Evans & Jörg Zinken (eds.), *The cognitive linguistics reader*, 834–848. London: Equinox.

Tomasello, Michael. 2007b. Cognitive linguistics and first language acquisition. In Dirk Geeraerts & Cuyckens Hubert (eds.), *The Oxford handbook of cognitive linguistics*, 1092–1112. Oxford: Oxford University Press.

Vogt, Paul & Elena Lieven. 2010. Verifying theories of language acquisition using computer models of language evolution. *Adaptive Behavior* 18(1). 21–35.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cllt-2020-0045).