The traceback method in child language acquisition research: identifying patterns in early speech*

STEFAN HARTMANN ⬥

University of Düsseldorf

NIKOLAS KOCH ⬥

LMU Munich

A N D

ANTJE ENDESFELDER QUICK ⬥

University of Leipzig

ABSTRACT

This paper discusses the traceback method, which has been the basis of some influential papers on first language acquisition. The method sets out to demonstrate that many or even all utterances in a test corpus (usually the last two sessions of recording) can be accounted for with the help of recurrent fixed strings (like *What's that?*) or frame-and-slot patterns (like [*What's* X?]) that can also be identified in the remaining dataset (i.e., the previous sessions of recording). This is taken as evidence that language learning is much more item-based than previously assumed. In the present paper we sketch the development of the method over the last two decades, and discuss its relation to usage-based theory, as well as the cognitive plausibility of its components, and we highlight both its potential and its limitations.

KEYWORDS: language acquisition, traceback, usage-based theory

---

1. Introduction

Over the last decades, usage-based approaches to language acquisition have become more and more influential in research on first language acquisition. In contrast to the long-dominant generative approach, which posits that language learners do not receive enough input to fully learn a language, and which therefore assumes an innate Universal Grammar (see, e.g., Valian, 2014, for discussion), usage-based approaches argue that children learn language based on the input they receive, making use of domain-general cognitive mechanisms like pattern finding and intention reading (see Tomasello & Lieven, 2008). Language acquisition is thus argued to be strongly ITEM-BASED, that is, organized around concrete, particular words and phrases (see, e.g., Tomasello, 2003). More specifically, usage-based approaches such as, e.g., Tomasello (2003) argue that children acquire language by learning fixed chunks (like *What's this?*) as well as schemas with an open slot, so-called frame-and-slot patterns (like [*What's* X?]). Tomasello (2000, p. 77) informally describes this as a "cut-and-paste" process. This is of course a simplifying metaphor as usage-based approaches also assume that children develop a fine-grained taxonomic network of form–meaning pairs – a 'constructicon' – when learning a language. What Tomasello's metaphor highlights, however, is that language in general, and child language in particular, is highly formulaic, which in turn provides the basis for learning and productively using language (see, e.g., Dąbrowska, 2014).

In this paper, we discuss one particular method that has been developed in the context of usage-based approaches to language acquisition. The traceback method sets out to test whether and to what extent children's utterances can be accounted for on the basis of a limited inventory of fixed chunks and frame-and-slot patterns. It has been adopted in a number of highly influential studies on first language acquisition (e.g., Lieven, Behrens, Speares, & Tomasello, 2003; Dąbrowska & Lieven, 2005; Lieven, Salomo, & Tomasello, 2009). The gist of the method is to account for as many utterances as possible in one part of the corpus by 'tracing them back' into the remaining dataset (see Section 2 for details). By doing so, the method aims at showing that even novel utterances can be accounted for by frame-and-slot patterns that already occur in previous utterances, which lends support to the theoretical assumption that language acquisition and language use are strongly item-based.

The aim of the present paper is to discuss the potential and limitations of the method in the light of usage-based theory. We offer a review of the most important published traceback studies so far, we discuss some criticisms that have been raised against the method, and we reflect on its potential and limitations. As such, our aim is to address two groups of readers: on the one hand, novices who want to gain a first overview of the method and its theoretical motivation; on the other hand, experienced researchers familiar with the

traceback method who want to critically reflect on the strengths and limitations of the method and who are potentially even interested in refining the method further (see Section 4 for some recently developed approaches building on the traceback method).

In the remainder of this paper, we first discuss which theoretical assumptions underlie the method, to what extent modifications took place over time, and how these modifications were theoretically motivated (Section 2). Our main argument will be that the connection between the traceback method and usage-based theory is twofold: on the one hand, the traceback method is used in a CONFIR-MATORY way to test key assumptions of usage-based approaches to language acquisition. On the other hand, it is used in an EXPLORATIVE way to assess how children acquire language, taking the pre-assumptions of the usage-based approach to language acquisition for granted. On the basis of this distinction, we discuss the potential and limitations of the method (Section 3). Section 4 summarizes our main arguments and points to avenues for further research.

2. Traceback: from theory to method and back again

2.1. THEORETICAL ASSUMPTIONS

The traceback method investigates how early multiword utterances of children can be composed of previously produced or perceived utterances. It assumes that, in acquiring language, children make use of fixed strings like *What's this?* on the one hand and frame-and-slot patterns like [*What's* THING?] on the other. This is rooted in key assumptions of usage-based linguistic theory.

'Usage-based theory' is a cover term for various theoretical approaches to language that share a number of key assumptions but also differ in important respects. While we cannot offer an in-depth discussion of usage-based approaches to language acquisition in general here (see, e.g., Behrens, 2009, for an overview), some remarks on the theoretical background of traceback studies are in order. Usage-based approaches assume that children learn language on the basis of the input they receive, not on the basis of an innate Universal Grammar. In particular, the processes of entrenchment, categorization, and schema formation are seen as crucial for language acquisition (see Behrens, 2009, p. 386; also see Schmid, 2020). Let us briefly discuss these concepts in turn: like many linguistic terms, ENTRENCHMENT can be interpreted both as a process and as a result state. As a process, it refers to the strengthening of the mental representation of a linguistic unit; as a state, it refers to the strength of representation of a linguistic unit (see, e.g., Taylor 2002, p. 590; Blumenthal-Dramé, 2012, p. 1). There is a broad consensus that the frequency with which a unit is encountered is a major determinant of entrenchment, although other factors (e.g., salience) seem to play a role as well and are increasingly emphasized in the current literature (see, e.g., Blumenthal-Dramé, 2012, and Schmid, 2017).

The concepts of CATEGORIZATION and SCHEMA FORMATION are closely connected: by detecting similarities and dissimilarities and by filtering out elements that do not recur (Behrens, 2009, p. 386), i.e., by forming categories, language users garner a large amount of abstract knowledge about the linguistic constructions they encounter. This is what Behrens (2009) calls 'schema formation', adopting Langacker's (1987, p. 492) definition of a schema as a "semantic, phonological, or symbolic structure that, relative to another representation of the same entity, is characterized with lesser specificity and detail". These generalizations in turn allow for the formation of what is often called 'constructional schemas', i.e., patterns at various levels of abstraction. For our purposes, the most important among those are constructions that are partially lexically filled and have one or more open slots, e.g., [*What's* X?]. In Construction Grammar, such patterns are usually called 'partially schematic' or 'partially filled constructions'. In the context of usage-based approaches to language acquisition, they are often referred to as 'frame-and-slot patterns', which is the term that we will use in the present paper. As, e.g., Goldberg (2006, p. 48) points out, there is a broad consensus that children store concrete exemplars but also form abstractions over these exemplars. There is some debate, however, whether children store (partially) abstract schemas such as frame-and-slot patterns or if they only make local and on-the-fly abstractions (see, e.g., Ambridge 2020a, 2020b). We will briefly return to this issue in Section 4.

Despite such differences between various usage-based approaches, they all have in common that they reject the nativist assumption that the input that children receive is too sparse for them to fully learn their native language ('poverty-of-stimulus argument'; see Berwick, Pietroski, Yankama, & Chomsky, 2011, for a recent defense). This is where the traceback method enters the picture. It was originally designed to demonstrate that virtually all utterances in early child language can be attributed to a relatively small set of units that are frequently attested.

2.2. IDENTIFYING FRAME-AND-SLOT PATTERNS AND FIXED STRINGS

The goal of the traceback method is "to account for all of the utterances produced in a short test corpus (e.g. the final session of recording) using lexically specific schemas and holophrases attributed to the child on the basis of previous sessions" (Ambridge & Lieven, 2011, p. 218). To this end, the corpus under investigation is first split into two parts, usually called MAIN CORPUS and TEST CORPUS (e.g., Lieven et al., 2009). The test corpus usually consists of the last one or two sessions of recording. The aim is now to 'trace back' the child's utterances in the test corpus – the so-called TARGET UTTERANCES – to precedents in the main corpus. These precedents are called COMPONENT

UNITS (e.g., Dąbrowska & Lieven 2005, p. 447). Two types of component units can be distinguished: on the one hand, FIXED STRINGS (also called '(fixed) chunks' or 'fixed phrases'), i.e., verbatim matches; on the other hand, FRAME-AND-SLOT PATTERNS (also called 'lexically specific schemas'). The traceback method does not assume more abstract categories because it is generally assumed that children start out with concrete and item-based constructions with only limited and local abstractions (see, e.g., Ibbotson & Tomasello, 2009, p. 60). Usage-based theories that assume stored abstractions and generalizations (unlike radically exemplar-based theories; see Sections 2.1 and 4) argue that those emerge gradually over the course of language acquisition (see, e.g., Abbot-Smith & Tomasello, 2006; Goldberg, 2006). This is why more abstract constructions that are often posited in constructionist approaches to (adult) language – such as the ditransitive construction or the caused-motion construction – are not assumed to play a role in early language acquisition. The traceback method therefore only works with two types of component units: fixed strings and frame-and-slot patterns.[1]

In order to find precedents, the most similar string is identified from the main corpus. If there is a verbatim match, it is considered a fixed string (e.g., *I want this one* in the toy example in Figure 1). If an utterance is not attested verbatim in the main corpus frequently enough (most traceback studies employ a frequency threshold of 2 occurrences in the main corpus), the analyst tries to derive it by searching for partial matches and 'constructing' the utterance with the help of predefined operations (see Section 2.2.1). This leads to the identification of frame-and-slot patterns, which represent the second type of component units. For instance, the second target utterance in Figure 1, *I want a paper*, does not have a verbatim precedent in the main corpus. However, we do find *I want a toy* and *I want a cake*, which allow for positing the frame-and-slot pattern [*I want a* THING]. In principle, however, the frame-and-slot pattern could also be [*I want* THING] or [*I want a* X], without a semantic specification of the slot filler. This shows that, while identifying verbatim matches is easy enough, positing frame-and-slot patterns entails numerous follow-up questions. This is why traceback studies differ considerably in how they operationalize the identification of frame-and-slot patterns. The box in the lower right-hand side of Figure 1 illustrates three 'parameters' (in a theory-neutral sense) along which the different applications of the traceback method

---

[1] Miorelli (2017), however, extends the method to allow for more flexible word order patterns in languages like Italian, which is the subject of his investigation. Apart from fixed chunks and frame-and-slot patterns, he posits "fully-specific packets" and "schematic packets" (Miorelli, 2017, p. 182), which account for units that occur in the main corpus with at least two different linear orders. For instance, the utterances *è caduta la tenda* 'the curtain has fallen' and *la torre è caduta* 'the tower has fallen' give rise to the schematic packet *La<THING + è<caduta* (Miorelli, 2017, p. 49).
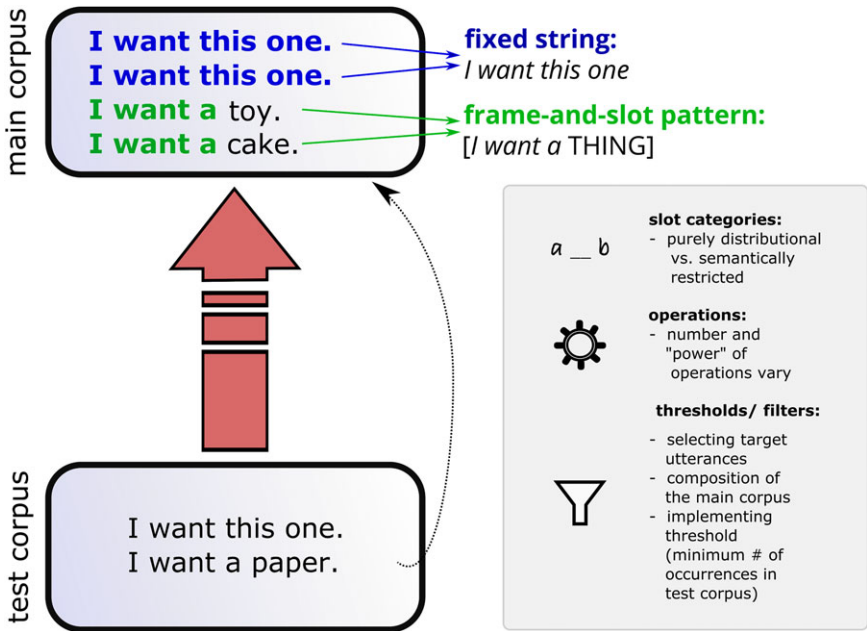
Fig. 1. The traceback method, illustrated with (constructed) examples from Dąbrowska & Lieven (2005). The box on the lower right-hand side summarizes the 'parameters' along which different applications of the method vary.

vary. These methodological choices will be the topic of the remainder of this section. We will focus on four influential papers: Lieven et al. (2003), Dąbrowska and Lieven (2005), Lieven et al. (2009), and Vogt and Lieven (2010).

The traceback method constrains the possibilities for positing frame-and-slot patterns by working with a finite set of operations. These operations will be discussed in more detail in Section 2.2.1. Furthermore, the question arises whether patterns should be identified purely on the basis of distributional information or if semantic aspects should be taken into account as well. Most traceback studies opt for the latter approach by positing semantic constraints for the open slots in frame-and-slot patterns. These will be the topic of Section 2.2.2. Section 2.2.3 discusses thresholds and 'filters' that decide which utterances enter the main and test corpus and which strings or patterns qualify as component units.

2.2.1. *Operations*

Usage-based linguistics assumes that children's language learning relies fundamentally on pattern finding and advanced categorization skills (see, e.g., Tomasello & Lieven, 2008; Ibbotson, 2020, Chapter 3). The traceback method is

TABLE 1. *Traceback operations (adopted from Koch, 2019, p. 180; constructed examples based on German data reported on there).*

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: [*I want **the** REF*] component unit: ***the** teddy* 'the teddy' → *I want **the** teddy*
SUBSTITUTE	a component unit is inserted into the open slot of a schema	schema: [*this is ATT*] component unit: *high* 'high' → *this is high*
ADD	linear juxtaposition of component units	component unit 1: *go away* component unit 2: *Mommy* → *go away Mommy!*

Note. Boldface indicates lexical overlap between two component units.

informed by what is known about such learning mechanisms, but it does not operationalize them directly. Instead, we can think of traceback as a way of reverse-engineering such cognitive processes: It aims to reconstruct the 'cut-and-paste' strategy that children follow in language acquisition in reverse by using different OPERATIONS to derive target utterances in the test corpus from their closest match in the main corpus. Nevertheless, the operations are often argued to represent actual mechanisms that children make use of when constructing new utterances. Note, however, that the authors of the different traceback studies take different positions regarding their cognitive plausibility: while Lieven et al. (2009, p. 502) make it very clear that "[w]e are not suggesting that the children's utterances are actually constructed by the operations given here, nor that the schemas and fixed strings that are identified are necessarily present in the child's linguistic representations", Dąbrowska and Lieven (2005, p. 442) assume that children's production of novel utterances involves the two operations that they use in their study. We will return to these issues in Sections 3 and 4.

The number and scope of the operations as well as their names differ between the individual traceback studies. In the remainder of this section, we will focus on the three operations that are used (under different names) across all traceback studies: SUBSTITUTE, SUPERIMPOSE, and ADD (see Table 1).

The operations SUBSTITUTE and SUPERIMPOSE both operationalize the idea of frame-and-slot patterns. As Tomasello (2003, p. 114) points out, at around 18 months of age, children start to use "pivot schemas" like [*more* X], e.g., *more milk*, *more juice*. These are characterized by one word or phrase (here: *more*) that basically serves as the 'anchor' of the utterance, while the linguistic items with which this pivot is combined fill the variable slot. Later on, these pivot schemas give rise to ever more complex constructional patterns that also consist

233

of fixed parts and variable slots. Many traceback studies use two different processes to do justice to the varying degrees of complexity of frame-and-slot patterns: "In a SUPERIMPOSE operation, the component unit placed in the slot overlaps with some lexical material of the schema, while in a SUBSTI-TUTE operation it just fills the slot" (Vogt & Lieven, 2010, p. 24). As an example for SUBSTITUTE, consider the target utterance *I want a paper* (from Dąbrowska & Lieven, 2005, p. 439). If utterances like *I want a toy* and *I want a banana* can be found in the main corpus, it is assumed that children substitute one unit for another, and the frame-and-slot pattern [*I want a* THING] can be posited. As an example for SUPERIMPOSE, consider the target utterance *I can't open it* (from Vogt & Lieven, 2010, p. 25). The whole chunk does not qualify as a component unit as it only occurs once in the main corpus – for a successful traceback, it would have to occur twice (see Section 2.2.3 on thresholds). However, strings like *I can't see it*, *I can't sing it*, and others occur in the main corpus, which allows for positing a frame-and-slot pattern. Also, the string [*can't open it*] occurs in the main corpus more than once. As the matching string *can't open it* shares material with the schema [*I can't* PROCESS *it*], SUPER-IMPOSE rather than SUBSTITUTE is used. Note that in Dąbrowska and Lieven (2005), the operation called SUPERIMPOSITION covers both substi-tution and superimposition, while Lieven et al. (2009) use only SUBSTITUTE but not SUPERIMPOSE. The application of SUPERIMPOSE and SUBSTI-TUTE is constrained by the semantic slot types that will be discussed in more detail in Section 2.2.2: They are allowed only when the filler has the semantic properties required by the slot (see Dąbrowska & Lieven, 2005, p. 448).

In addition to these two operations, there is also the possibility of linear juxtaposition. However, SUBSTITUTE and SUPERIMPOSE are always used first. Vogt and Lieven (2010, p. 24) motivate this methodological choice with the central role of frame-and-slot patterns in the usage-based approach.

The operation that allows for the linear juxtaposition of strings is called ADD or JUXTAPOSE. While Lieven et al. (2003) made no restrictions for this operation, the subsequent studies allowed it only if the combination is syntactically as well as semantically possible in any order. This means, for example, that conjunctions such as *and* or *because* could not be used with the ADD operation (Vogt & Lieven, 2010, p. 24): For example, in a hypothetical utterance like *And I want ice cream*, a juxtaposition of the component units *and* + [*I want* THING] would not be possible because the reverse order is ungrammatical (*I want ice cream and.*) This essentially limits the application of ADD to vocatives like *mommy* or adverbials like *now* and *then*. The main motivation for this constraint is that, otherwise, the ADD operation might yield implausible derivations (see Dąbrowska & Lieven, 2005, p. 439). How-ever, Kol, Nir, and Wintner (2014, p. 194) question the cognitive plausibility of the restriction, stating that it is unclear how the child should know this

constraint. In addition, they point out that it does not become clear how the positional flexibility of a given unit is determined (i.e., whether the unit has to be attested in both positions in the corpus or whether the native-speaker intuitions of the annotators are decisive).

While the use of these three derivations (and the principle of always using SUBSTITUTE and SUPERIMPOSE before ADD) allows for a systematic and principled way of deriving target utterances, there are still many cases where more than one derivation is possible: For example, Dąbrowska and Lieven (2005, pp. 459f.) discuss the example *I want some orange juice*. Their main corpus allows for positing [*I want some* X] as well as [*I want* X] as component units. Therefore, the following principles are posited: (1) the largest possible schemas are used; (2) the slots are filled by the longest available units; and (3) the minimum number of operations is taken (Lieven et al., 2009, p. 489). In our example, then, [*I want some* X] would prevail over [*I want* X] as it is the largest possible schema.

The underlying assumption is that in the production of utterances several paths are activated in parallel to form the utterance "and that the simplest one wins the race" (Dąbrowska & Lieven, 2005, p. 460). Dąbrowska and Lieven make the (of course idealizing) assumption that larger units are always 'simpler' and are therefore preferred by speakers when constructing new utterances: the more general units – in our example, [*I want some* X] – are cognitively more demanding even though they may be more entrenched as we encounter them more frequently. Similar assumptions can be found elsewhere in the literature, not just in research on (first) language acquisition. For example, Sinclair's (1991, p. 110) idiom principle posits that speakers have available "a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments".[2] He states "a natural tendency to economy of effort" as one potential explanation. Variations of this idea are widespread in the usage-based literature (cf. the concept of 'prefabricated units' or 'prefabs'; e.g., Bybee 2007). From this perspective, it makes sense to assume that units that frequently occur together are stored as chunks – or, as Bybee (2007, p. 316) puts it, "items that are used together fuse together". Wherever a fixed chunk is available, it can be expected that it is preferred to a 're-construction' of the same unit from its component units, which again lends support to the methodological decision of always choosing the longest units.

2.2.2. *Semantic slot categories*
The first traceback study (Lieven et al., 2003) took a purely distributional, string-based, and strictly bottom-up perspective that did not take semantics

---

[2] Thanks to an anonymous reviewer for pointing us to this reference.

TABLE 2. *Types of slots (from Vogt & Lieven, 2010).*

Type of slot	Example utterances		Frame-and-slot pattern
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.	
POSSESSOR	INV	This is my **favorite**.	[POSSESSOR *favorite*]
	MOT	Yeah it's your **favorite** that one, isn't it?	
UTTERANCE	CHI	Open the door **Mummy**.	[UTTERANCE *Mummy*]
	INV	There's the cook **Mummy**.	

Note. Boldface indicates lexical overlap between two component units.

into account (see Lieven et al., 2009, p. 484). In later studies, however, Lieven and colleagues proposed different slot categories within frame-and-slot patterns based on semantic relations: in an example like [*I wanna have the* X], a REFERENT slot would be assumed. Table 2 provides an overview of all slot types posited in the literature.

These slot types are based on the assumption that children can make semantic generalizations about the content of these slots from early on. This assumption is rooted in usage-based theory, with some key proponents of a usage-based approach to language acquisition explicitly relating it to Langacker's (1987, 1991) Cognitive Grammar (e.g., Dąbrowska & Lieven, 2005; Abbot-Smith & Tomasello, 2006). However, the assumptions are not purely theoretical: as Dąbrowska & Lieven (2005, p. 444) point out, there is good evidence in the developmental literature that children start categorizing objects even before they can speak – for example, a variety of early experimental approaches summarized by Mandler (1992) have looked into children's concepts of animacy and agency by assessing their behavior when seeing, for example, a hand vs. a block of wood picking up an object. Also, Dąbrowska and Lieven (2005, p. 444) refer to tentative corpus-based evidence that children begin to form semantic subclasses of verbs from around their second birthday (Pine, Lieven, & Rowland, 1998). Abbot-Smith and Tomasello (2006, p. 283) point out that children are able to distinguish utterances referring to events, processes, and states fairly early on based on relational similarity. Recent studies on passive constructions (that can, however, potentially be generalized

to other constructions as well) have lent further support to the hypothesis that children form 'semantic construction prototypes' in the sense that their representations of constructions are both abstract and semantically constrained (Bidgood, Pine, Rowland, & Ambridge, 2020).

2.2.3. *Thresholds and filters*

In this section, we address two related questions that are crucial for any traceback study: Which utterances are taken into account in the test corpus and in the main corpus, and when do units found in the main corpus qualify as component units? The answers to these questions depend on: (a) the division into test corpus and main corpus; (b) the selection of utterances in the main corpus and test corpus that are actually taken into account in the traceback procedure; and (c) the criteria for establishing component units. We will discuss these aspects in turn.

As for (a), it would be possible, in principle, to abandon the rigid distinction between test and main corpus and to trace each utterance in the entire corpus back to all previous utterances (as in Quick, Hartmann, Backus, & Lieven, forthcoming), but this entails the problem that the 'pool' of preceding utterances grows with each utterance, which means that it may be more likely to find precedents for later utterances than for earlier ones. The division into main and test corpus ensures that the set of utterances to which the target utterances are traced back remains consistent. Note, however, that Lieven et al. (2003, p. 338), Dąbrowska and Lieven (2005, p. 448), and Lieven et al. (2009, p. 489) also use an additional criterion for acknowledging a string as a component unit: if a fixed string that occurred in a target utterance occurred in the five immediately preceding utterances in the test corpus, then it is also seen as a component unit even if it does not occur at all in the main corpus. In Dąbrowska and Lieven (2005, p. 448) and Lieven et al. (2009, p. 489), this special rule is, however, restricted to single words.

This leads us to (b), i.e., to the question of which utterances are taken into account in the test and main corpus. As for the test corpus, usually all the child's multiword utterances are taken into account, i.e., one-word utterances are not included in the set of target utterances.[3] Dąbrowska and Lieven, however, focus on question constructions, which is why they take only syntactic questions into account as target utterances, i.e., "utterances involving either a preposed auxiliary and a subject (for yes/no questions) or a preposed WH-word and at least one other word" (Dąbrowska & Lieven, 2005, p. 440).

---

[3] Koch (2019) uses a narrower definition of 'multi-word utterances' by distinguishing them from both one-word and two-word utterances. In his traceback study, he therefore only takes utterances with three or more words into account.

This is a major difference between Dąbrowska and Lieven's study and all other traceback studies reviewed in this paper, but the methodological approach is the same as in subsequent traceback studies. Also note that this restriction to *wh*-questions only applies to the set of target utterances. To account for the target utterances, all utterances in the main corpus (not just question constructions) are taken into account.

As for the main corpus, a key issue is the methodological choice of including or excluding the caregivers' input. This aspect has been discussed extensively, but it has also been shown that including or excluding the input does not drastically affect the results (see, e.g., Dąbrowska & Lieven, 2005; Koch, Hartmann, & Endesfelder Quick, forthcoming). The theoretical motivation for including the caregivers' input is that usage-based approaches assume that children's linguistic knowledge is shaped by the input they receive, which has also been substantiated in a number of studies (e.g., Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Behrens, 2006). In Vogt and Lieven (2010), taking the input into account is motivated by their research question: they compare the results of a traceback study to those of usage-based computational models, which implement the assumption that children learn language based on their input – hence, taking the input into account makes the traceback study "closer in spirit" to these models (Vogt & Lieven, 2010, p. 23). However, leaving out the input strengthens the case that the children actually have the identified units available in one way or another (for a discussion of cognitive plausibility, see Section 3): while it seems plausible to assume that a child will be familiar with a linguistic unit that is repeatedly used in the caregivers' input, we can be much more confident that it is an entrenched unit if it is actually used by the child – even more so if it is used repeatedly. This leads us to (c), the criteria for identifying component units, i.e., for assuming that the fixed strings or the patterns identified are actually available to the child. The most important aspect here is the frequency threshold. Most traceback studies work with a frequency threshold of two occurrences in the main corpus.

As Dąbrowska (2014, p. 621) points out, a threshold of two occurrences may seem very low at first glance, but it should be kept in mind that corpora only capture a small proportion of the child's linguistic experience; as for the children's utterances, Lieven et al. (2009, p. 493) estimate that their sample captures 7–10% of what the children say. Given these numbers, the actual frequencies of the fixed strings and frame-and-slot patterns detected by the traceback method in the child's linguistic experience is expected to be much higher (see Dąbrowska 2014, p. 640).

Another aspect that plays a role when it comes to identifying component units are coding choices. In most traceback studies so far, the fixed chunks and frame-and-slot patterns were identified via (semi-)manual data annotation. Especially the identification of semantic slot categories (see Section 2.2.2) often

238

requires annotation choices that become all the more important the more slot categories are used. Later traceback studies like Kol et al. (2014) and Koch (2019) have also used fully automatic procedures, using the morphological and syntactic annotations – especially the POS tagging – available in their corpora as proxies for identifying semantic slot categories. This will be discussed in more detail in Section 3.2.

Summing up, the traceback method relies on a variety of choices that the researcher has to make. They are guided by the research question of each individual study as well as by theoretical pre-assumptions. By and large, however, they are largely arbitrary decisions required by the method, and various studies have already investigated to what extent these methodological choices influence the results of traceback studies (e.g., Dąbrowska & Lieven, 2005; Dąbrowska, 2014; Koch et al., forthcoming).

2.3. THE TRACEBACK METHOD IN ACTION

In this section, we give a brief overview of the main results of published traceback studies, focusing on overarching tendencies in research on monolingual L1 acquisition. For an exhaustive (to the best of our knowledge) overview of the design and results of all published traceback studies so far, see Table 3 in the Appendix.[4]

As discussed in Section 2.2, the concrete operationalization of the basic idea of tracing back utterances to potential precedents differs across different studies. The methodological choices of course have an effect on the traceback results, but they have shown to be relatively minor (e.g., Dąbrowska & Lieven, 2005; Dąbrowska, 2014; Koch et al., forthcoming). All studies have consistently shown that a high proportion of target utterances can be successfully traced back. In addition, the traceback results have served as the starting point for a more detailed analysis of the patterns that have been identified, which can in turn be related to assumptions about the psychological reality of these units. For example, Dąbrowska and Lieven (2005) use their traceback results to assess how the child's linguistic abilities change between the ages of two and three. In order to do this, they compare the operations and slot types at age 2;0 and 3;0, showing that "[t]he three-year-olds' output is less stereotypical and repetitive in that they superimpose over a wider range of slots and are able to apply a larger number of operations per utterance" (Dąbrowska & Lieven, 2005, p. 456). In addition, Bannard and Lieven (2009) discuss the results of the

---

[4] There are a few traceback studies on multilingual data (e.g., Quick et al., forthcoming) that are not mentioned in Table 3 as their methodology deviates quite significantly from that of the 'classic' traceback studies reviewed here. Also, related methods that build on traceback are not taken into account (e.g., Bannard et al.'s (2009) traceforward; McCauley & Christiansen's (2017) chunk-based learner).

'classic' traceback studies in connection with experimental evidence on multi-word storage.

As mentioned in Section 2.2.3, the quantitative – and more confirmatory – results are often accompanied by more exploratory analyses. For example, the number of operations required for traceback is assessed and correlated with the children's mean length of utterance (MLU) by Lieven et al. (2009) for English, Miorelli (2017) for Italian, and Koch (2019) for German. As expected from a usage-based perspective, the number of exact matches decreases with increasing MLU, while the proportion of tracebacks that require multiple operations increases. In addition, Dąbrowska and Lieven (2005), Lieven et al. (2009), and Koch (2019) also offer a more detailed analysis of the semantic slot categories. Lieven et al. (2009, p. 494) show that the vast majority of slot fillers belongs to the REFERENT category but that the proportion of other slot fillers (most notably PROCESS) increases with increasing MLU. Other traceback studies arrive at similar results, and Koch (2019, p. 223) shows that the general tendency also holds for German data.

The final step of a traceback study often consists in an analysis of the instances that could not be derived to the main corpus, the so-called fails. Usually, researchers distinguish between "lexical fails" and "syntactic fails" (e.g., Dąbrowska & Liven, 2005, pp. 453–455; Lieven et al., 2009, pp. 492–494; Koch, 2019, pp. 245–254). In the former case, a word occurs in the target utterance that is not attested in the main corpus with sufficient frequency; in the latter case, no relevant component units for a novel utterance can be found (see Lieven et al., 2009, p. 489). The distinction between syntactic and lexical fails is important from a theoretical perspective: as Dąbrowska and Lieven (2005, p. 453) point out, "the fact that the child used a word constitutes reliable evidence that he/she knows it". Even proponents of a nativist approach to language acquisition would of course never claim that children come equipped with innate lexical items. Thus, we must assume that the child has previously encountered the word in a context that wasn't sampled. Syntactic fails, by contrast, provide a greater challenge when looking at the results from a confirmatory perspective, but they also allow for detecting interesting patterns from an explorative perspective. For example, Dąbrowska and Lieven (2005, p. 453), analyzing the syntactic fails in their data, observe that "a very high proportion (62 percent) of the problematic utterances are ill-formed by adult standards". Therefore, they conclude that, in these utterances, children experiment by going beyond what they already know, rather than applying rules they have already mastered (Dąbrowska & Lieven, 2005, p. 455).

Analyzing the syntactic fails can also point to individual differences between the children under investigation. For example, in Koch's (2019) study of four monolingual German children, the proportion of syntactic fails varies between 5% and 13%. Most of the fails occurred in the corpus of the child who was also

240

the most advanced language learner, according to other measures such as MLU (also see Koch et al., forthcoming). This is not surprising: if a child's language use is highly creative and productive, the overall traceback success will be low compared to a child whose speech is highly formulaic. However, an in-depth qualitative analysis of the syntactic fails can point to individual differences in creative language use – for example, the derivations could fail because the children use constructions that are considered ungrammatical (and therefore do not occur in the child's input, which is usually included in the main corpus; see Section 2.2.3), or because the child uses highly complex utterances that cannot be accounted for using only the simple operations employed by the traceback method. Thus, differences in how many utterances can be traced back successfully, as well as differences in the number of operations required for the successful tracebacks, can give clues to different styles of language learning and use.

3. Potential and limitations of the traceback method

As we have argued throughout this paper, the traceback method can be used in a confirmatory and in an exploratory way. We will now address each of these perspectives in turn: from the confirmatory perspective, the crucial question is whether the results of traceback studies actually lend support to the key hypotheses of usage-based theory (Section 3.1). From an exploratory perspective, the question arises to what extent traceback results allow for drawing conclusions about language learning and/or the cognitive processes underlying child language acquisition. This is closely connected to the question of cognitive plausibility that we have already addressed above. In Section 3.2, we return to this issue. As in many corpus-based approaches, confirmatory and explorative aspects can of course not always be neatly teased apart, as quantitative and qualitative analyses are closely connected and blend into one another. Apart from the overarching hypothesis that, due to the item-based nature of language learning, a large proportion of target utterances can be traced back successfully, the traceback method is also used to investigate much more specific hypotheses (see, e.g., Koch, 2019, pp. 154–156), e.g., relating to the relative frequency of different operations or slot types across different age groups. Note, however, that addressing these more specific hypotheses already presupposes the specific operationalization offered by the traceback method. This means that if the method itself should turn out to yield results that are cognitively implausible, this would also cast doubt on the results obtained with regard to these lower-level research questions. In the remainder of this section, we will, however, argue that there is no reason to doubt that the traceback results are – at least to some extent – cognitively plausible, although we will also argue in Section 4 that some aspects of the method can potentially be refined.

3.1. THE MEANING OF SUCCESS: WHAT SUCCESSFUL TRACEBACKS CAN TELL US

The traceback studies have consistently shown that a large proportion of utterances in the test corpora can be accounted for with fixed strings and frame-and-slot patterns that can be found in the main corpus. This has been taken to lend support to the hypothesis that "what children say is closely related to what they have said previously, not only because of extraneous factors […] but precisely because this is how they build up their grammars" (Lieven et al., 2009, p. 502). However, the method has also been criticized for being too unconstrained. For example, Kol et al. (2014) showed that even tracing back target utterances with reversed (*you can do it the other way > way other the it do can you*) or randomized word order (*other you can way it do the*; see Kol et al., 2014, pp. 192–195) led to an almost identical proportion of successful tracebacks compared to the original utterances (see Koch et al., forthcoming, for a more in-depth discussion). Given these results, Kol et al. (2014) argue that the method suffers from an overgeneralization problem in the sense that it can derive virtually any target utterance. In other words, the method is, in their view, still too unconstrained – a criticism already raised by Dąbrowska and Lieven (2005) against the first application of the method in Lieven et al. (2003). However, Kol et al.'s (2014) application of the method is not directly comparable to the classic traceback studies as there are major differences in their operationalization, as shown in Koch (2019) and Koch et al. (forthcoming). Also, the fact that ungrammatical utterances can be derived is not necessarily a problem. Quite to the contrary, if ungrammatical structures do occur in the target utterances, a successful derivation can help explain how they came about (see Koch et al., forthcoming, for a more in-depth discussion).

Despite these criticisms, the fact that the traceback success remained relatively high for reversed and randomized target utterances (even in replication studies using the original methodology, see Koch et al., forthcoming) shows that Kol et al. (2014) do have a point: if the traceback success is consistently very high, this might be due to other factors such as simple laws of word frequency distributions. As is well known, individual words as well as *n*-grams follow a Zipfian distribution (see, e.g., Yang, 2016, pp. 18f.; also see, e.g., Ellis, O'Donnell, & Römer, 2013). This means that a small number of words and word combinations occurs with a very high frequency while the bulk of words and *n*-grams is exceedingly rare. Thus, it can be expected that a relatively large proportion of all utterances in a small subset of any corpus will have equivalents in the remaining dataset (as has been shown by Dąbrowska, 2014, for adult language). On the other hand, given what Yang (2016, p. 18) calls "Zipf's long tail", it would of course be unreasonable to expect that A L L utterances in a test corpus can be successfully derived (see Koch, 2019, p. 265). After all, we

242

are just dealing with a very small sample of the child's linguistic experience, as has been pointed out repeatedly in different traceback studies (e.g., Dąbrowska & Lieven, 2005, p. 453; Dąbrowska, 2014, p. 621). These limitations of course do not invalidate the traceback method, but they have to be kept in mind when interpreting the results of traceback studies, and especially when comparing the results of different traceback studies: in the latter case, it has to be critically evaluated whether the differences actually reflect differences in language acquisition or if they can be explained by other (methodological) factors such as the size of main and test corpus (see Koch et al., forthcoming, for manipulation studies that test the influence of these and other variables).

Note that the results of traceback studies are, in principle, also compatible with other theoretical approaches – partly, however, because some usage-based concepts have also been adopted in frameworks that were previously perceived as being strongly opposed to the foundational concepts of usage-based theory. Perhaps most prominently, Yang (e.g., 2016), while sticking to the generativist assumption that there is an innate 'core' to language, relegates much of the explanatory burden to data-driven inductive learning. This approach also assumes that early child language is strongly item-based and would therefore predict that a large proportion of target utterances can be traced back successfully.

In sum, then, the high proportion of successful tracebacks that has been a consistent finding across all traceback studies is unsurprising given what we know about the skewed frequency distributions that are characteristic of language. However, as already pointed out above, assessing the number of successful tracebacks is only the starting point of traceback analyses, which leads us to the question of how to interpret traceback results.

3.2. FROM CORPUS TO COGNITION: INTERPRETING TRACEBACK RESULTS

Given the considerations about the confirmatory aspects of the traceback method, it is not surprising that most traceback papers focus on the more explorative perspective by offering a detailed qualitative analysis of the traceback results. The validity of any such interpretation of course partly depends on the question of whether the operations and slot categories used by the method can be connected to cognitive mechanisms underlying language learning, as discussed in Section 2.2. In other words, we are necessarily faced with the question whether and to what extent the assumptions of the method and the way they are operationalized can be considered cognitively plausible.

This can be illustrated by returning to the semantic slot categories discussed in Section 2.2.2. As we have seen, various authors differ in the constraints that they posit for open slots in constructions like [*I want* X] or [*What's* X

243

doing?]. Dąbrowska (2000) describes early frame-and-slot patterns as constructional schemas. Drawing on Langacker (1987, 1991), she assumes abstract representations with a specific semantic specification for the open slots, e.g., [ANIMATE *WANT* THING].

From a methodological perspective, the question arises how the semantic slot types are operationalized. The algorithms used by Kol et al. (2014) and Koch (2019) use morphological and syntactic information as proxies for semantic categories. Especially part-of-speech categories play a crucial role in both implementations. This builds on the core idea of cognitive-linguistic and usage-based approaches that word class categories "are semantically definable not just at the prototype level, but also schematically, for all category members" (Langacker, 2005, p. 121) Importantly, there is also ample evidence that children can distinguish the types of entities that the major word classes denote at a very early age, as has been shown time and again in the literature since Brown's (1957) seminal paper (see, e.g., Bloom, 2000, for discussion). As Abbot-Smith and Tomasello (2006, p. 283) point out, children are capable of detecting whether an utterance refers to an event, a process, or a state very early on (see also Tomasello & Brooks, 1999). Thus, it seems plausible to assume that such semantic aspects form part of their knowledge about individual frame-and-slot patterns from early on.

Another aspect where the question of cognitive plausibility plays a major role in the interpretation of traceback results is the issue of entrenchment and its relation to frequency (see, e.g., Blumenthal-Dramé, 2012, and Schmid, 2017, for discussion). As we have seen throughout this paper, the methodological choices of the traceback method are often motivated by theoretical assumptions, some of which are discussed controversially. For example, a basic assumption is that children abstract away frame-and-slot patterns from usage events and then have them available as stored abstractions. However, the degree to which this is actually plausible is still subject to considerable discussion (see Ambridge, 2020a, and the response papers to this, e.g., Lieven, Ferry, Theakston, & Twomey, 2020, Zettersten, Schonberg, & Lupyan, 2020, as well as Ambridge, 2020b).

While the exact nature of the units that language users store is debated in usage-based approaches, there is broad agreement that one factor that strongly influences the degree of entrenchment of a unit is frequency. The traceback method usually operationalizes frequency in terms of an arbitrary threshold, as it is primarily interested in showing that speakers, paraphrasing Dąbrowska (2014), "recycle utterances".

However, if we are also interested in whether the patterns identified by the traceback method are actually cognitively entrenched patterns, it makes sense to (a) ask the question when exactly a unit can be counted as entrenched; and (b) take frequency into account in a more fine-grained way, as suggested by Kol et al. (2014). We will return to these issues in Section 4.

Summing up, the issue of entrenchment and frequency effects is closely connected to the potential and limitations of the traceback method in general. Traceback studies usually do not make use of fine-grained measures of the chunks and frame-and-slot patterns that are identified. Instead, they work with arbitrary frequency thresholds, which is a result of the fact that the method was originally conceived to demonstrate the ubiquity of formulaic patterns and not necessarily as a way of exploring them in more detail. As subsequent studies have gone further in their interpretation of traceback results, the question of how to take frequency effects into account has become more relevant. In addition, as language learning always takes place in specific contexts, the questions that become more relevant are whether and to what extent different settings and environments influence the frequency with which a child encounters particular linguistic units, and how such contextual factors may influence entrenchment.

4. Discussion and conclusion

In this paper, we have discussed the potential and limitations of the traceback method, which has been used in a number of highly influential papers on first language acquisition. While the concrete operationalization of the basic premise differs across the studies reviewed here, the traceback studies have consistently shown that a large proportion of children's utterances has precedents in utterances that the child has said or used before. But beyond that, the method can arguably also help investigate H O W constructions are acquired. Firstly, it can help provide an answer to the question of which patterns that a child uses can be considered entrenched. That said, it should be noted that a more detailed assessment of this question would require a more fine-grained operationalization of frequency effects. Secondly, analyzing the semantic slot categories in the frame-and-slot patterns identified by the method or taking a closer look at failed derivations can point to individual differences between children.

We have also addressed the issue of cognitive plausibility. In usage-based approaches, taking corpus data at face value has come to be widely regarded as a fallacy (see, e.g., Arppe, Gilquin, Glynn, Hilpert, & Zeschel, 2010). As Dąbrowska (2016, p. 486) points out, the patterns found in corpus data often do not reflect the way language is represented and organized in speakers' minds. These limitations of course do not just apply to the traceback method but to any corpus-based methodology. As Dąbrowska and Lieven (2005, p. 438) put it, "naturalistic data can only be indicative." For evaluating hypotheses about the cognitive underpinnings of language acquisition, converging evidence from various methods – e.g., corpus linguistics, behavioral experiments, and computational modelling – is needed. Note that the boundaries between these methods are arguably more blurred and permeable in language acquisition research than in any other domain of the language

245

sciences – for instance, data gathered via experimental elicitation methods are often investigated using corpus-linguistic methods, and corpus data serve as input for computational modelling. For example, in Vogt and Lieven (2010), the traceback results are used as input for a computational iterated learning model, and McCauley and Christiansen's (2017) "Chunk-Based Learner" aims to provide "a computationally explicit approach to the Traceback method" (McCauley & Christiansen, 2017, p. 649).

Returning to the question of cognitive plausibility, the very assumption of frame-and-slot patterns that is key to the traceback method is sometimes considered to be cognitively implausible. This issue is discussed extensively by Ambridge (2020a), who argues that the assumption of lexically based schemas like [He's ACTIONing it] only makes sense if they are taken as a metaphor for on-the-fly analogical generalizations. While, e.g., Abbot-Smith and Tomasello (2006) advocate a hybrid model that assumes both exemplars and stored abstractions, Ambridge himself argues for a radically exemplar-based account that does away with abstractions like the frame-and-slot patterns assumed in most traceback studies.

However, the traceback studies reviewed in this paper are actually compatible with both prototype-based and exemplar-based accounts: On a radically exemplar-based view (e.g., Ambridge, 2020a; see, e.g., Divjak & Arppe, 2013, and Schäfer, 2016, for a discussion of prototype-based and exemplar-based approaches from a linguistic perspective), frame-and-slot patterns can be seen as the generalizations that researchers make over many on-the-fly analogies on the basis of individual exemplars. In this view, they are a helpful heuristic tool for grouping observations that are frequently encountered in the data, which in turn point to generalizations that are frequently made by language users. On the more mainstream view that does assume abstractions, these generalizations are made not only by the researchers but also by the speakers themselves.

Summing up, the traceback method has yielded a variety of relevant insights into the role of fixed patterns and partially filled constructions in early child language. We have also shown that the method has evolved over time, and it continues to evolve, partly in response to the limitations that have been pointed out both in the 'classic' traceback papers and in Kol et al.'s (2014) critical evaluation. As mentioned above, one desideratum is to take frequency into account in a more fine-grained way. Various methods are conceivable to achieve this: for instance, a bottom-up pattern mining approach based on n-grams and skip-grams[5] could be used, following, e.g., the 'construction mining' approach pioneered by Forsberg et al. (2014) to detect constructions

---

[5] n-grams are chunks of n subsequent words: A string like I want a banana now could be split into the 3-grams I want a, want a banana, a banana now. Skip-grams allow for skipping a predefined number of words, e.g., skipping one word: I ___ a, ___ a banana, I want ___ , etc.

246

in adult language. In addition, McCauley and Christiansen's (2017) 'chunk-based learner', a computational-model-based extension of the traceback method (see McCauley & Christian, 2017, p. 639) already takes frequencies as well as transition probabilities between words into account.

Other recent extensions of the traceback method relate to the type of data taken into account: firstly, Dąbrowska (2014) has used the method to investigate the role of formulaic language in adult language. Secondly, Quick et al. (forthcoming) have adopted it to investigate code-mixing in bilingual language acquisition, using the code-mixed data in their dataset as the test corpus and the remaining data as the main corpus. Thus, the traceback method can be seen as a pioneering approach in a family of usage-based and constructional methods that are still being constantly refined, making use of the ever more sophisticated computational techniques available now.

REFERENCES

Abbot-Smith, K. & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review* **23**(3), 275–290.

Ambridge, B. (2020a). Against stored abstractions: a radical exemplar model of language acquisition. *First Language* **40**(5/6), 509–559.

Ambridge, B. (2020b). Abstractions made of exemplars or 'You're all right, and I've changed my mind': response to commentators. *First Language* **40**(5/6), 640–659.

Ambridge, B. & Lieven, E. (2011). *Child language acquisition: contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Arppe, A., Gilquin, G., Glynn, D., Hilpert, M. & Zeschel, A. (2010). Cognitive corpus linguistics: five points of debate on current theory and methodology. *Corpora* **5**(1), 1–27.

Bannard, C. & Lieven, E. (2009). Repetition and reuse in child language learning. In R. Corrigan, E. Moravcsik, H. Ouali & K. Wheatley (eds.), *Formulaic language: volume II: acquisition, loss, psychological reality, functional explanations*. Amsterdam: John Benjamins.

Bannard, C., Lieven, E. & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences* **106**(41), 17284–17289.

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

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics* **47**(2), 383–411.

Berwick, R. C., Pietroski, P., Yankama, B. & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science* **35**(7), 1207–1242.

Bidgood, A., Pine, J. M., Rowland, C. F. & Ambridge, B. (2020). Syntactic representations are both abstract and semantically constrained. *Cognitive Science* **44**(9), e12892.

Bloom, P. (2000). *How children learn the meanings of words*. MIT Press.

Blumenthal-Dramé, A. (2012). *Entrenchment in usage-based theories: what corpus data do and do not reveal about the mind*. Berlin: De Gruyter.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology* **55**(1), 1–5.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Dąbrowska, E. (2000). From formula to schema: the acquisition of English questions. *Cognitive Linguistics* **11**(1/2), 83–102.

247

Dąbrowska, E. (2014). Recycling utterances: a speaker's guide to sentence processing. *Cognitive Linguistics* **25**(4), 615–653.

Dąbrowska, E. (2016). Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* **27**(4), 479–491.

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

Divjak, D. & Arppe, A. (2013). Extracting prototypes from exemplars: What can corpus data tell us about concept representation? *Cognitive Linguistics* **24**(2), 221–274.

Ellis, N. C., O'Donnell, M. B. & Römer, U. (2013). Usage-based language: investigating the latent structures that underpin acquisition. *Language Learning* **63**, 25–51.

Forsberg, M., Johansson, R., Bäckström, L., Borin, L., Lyngfelt, B., Olofsson, J. & Prentice, J. (2014). From construction candidates to constructicon entries. *Constructions and Frames* **6**(1), 114–135.

Goldberg, A. E. (2006). *Constructions at work*. Oxford: Oxford University Press.

Huttenlocher, J., Vasilyeva, M., Cymerman, E. & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology* **45**(3), 337–374.

Ibbotson, P. (2020). *What it takes to talk: exploring developmental cognitive linguistics*. Berlin: De Gruyter.

Ibbotson, P. & Tomasello, M. (2009). Prototype constructions in early language acquisition. *Language and Cognition* **1**(**1**), 59–85.

Koch, N. (2019). *Schemata im Erstspracherwerb. Eine Traceback-Studie für das Deutsche*. Berlin: De Gruyter.

Koch, N., Hartmann, S. & Endesfelder Quick, A. (forthcoming). The traceback method and the early constructicon: theoretical and methodological considerations. *Corpus Linguistics and Linguistic Theory*.

Kol, S., Nir, B. & Wintner, S. (2014). Computational evaluation of the Traceback Method. *Journal of Child Language* **41**(1), 176–199.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Vol. 1: theoretical prerequisites*. Stanford, CA:Stanford University Press.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Vol. 2: descriptive application*. Stanford, CA: Stanford University Press.

Langacker, R. W. (2005). Construction Grammars: cognitive, radical, and less so. In F. J. de Ruiz Mendoza Ibáñez & M. S. Peña Cervel (eds.), *Cognitive linguistics* (pp. 101–159). Berlin: De Gruyter.

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

Lieven, E., Ferry, A., Theakston, A. & Twomey, K. E. (2020). Similarity, analogy and development in radical exemplar theory: a commentary on Ambridge (2020). *First Language* **40**(5/6), 600–603.

Lieven, E., Salomo, D. & Tomasello, M. (2009). Two-year-old children's production of multi-word utterances: a usage-based analysis. *Cognitive Linguistics* **20**, 481–507.

Mandler, J. M. (1992). How to build a baby II: conceptual primitives. *Psychological Review* **99** (4), 587–604.

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

Miorelli, L. (2017). *The development of morpho-syntactic competence in Italian-speaking children: a usage-based approach*. Unpublished doctoral dissertation, Northumbria University.

Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics* **36**(4), 807–830.

Quick, A. E., Hartmann, S., Backus, A. & Lieven, E. (forthcoming). Entrenchment and productivity: the role of input in the code-mixing of a German–English bilingual child. *Applied Linguistics Review*.

Schäfer, R. (2016). Prototype-driven alternations: the case of German weak nouns. *Corpus Linguistics and Linguistic Theory* **15**(2), 383–417.

Schmid, H.-J. (2017). A framework for understanding entrenchment and its psychological foundations. In H.-J. Schmid (ed.), *Entrenchment and the psychology of language learning. how we reorganize and adapt linguistic knowledge* (pp. 9–39). Boston, MA: De Gruyter.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Schmid, H.-J. (2020). *The dynamics of the linguistic system: usage, conventionalization, and entrenchment*. Oxford: Oxford University.

Taylor, J. R. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* **11**(1/2), 61–82.

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

Tomasello, M. & Brooks, P. J. (1999). Early syntactic development: a Construction Grammar approach. In M. Barrett (ed.), *The development of language* (pp. 161–190). Hove: Psychology Press.

Tomasello, M. & Lieven, E. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. J. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 168–196). London: Routledge.

Valian, V. (2014). Arguing about innateness. *Journal of Child Language* **41(S1)**, 78–92.

Vogt, P. & Lieven, E. (2010). Verifying theories of language acquisition using computer models of language evolution. *Adaptive Behavior* **18**(1), 21–35. doi 10.1177/1059712309350970

Yang, C. D. (2016). *The price of linguistic productivity: how children learn to break the rules of language*. Cambridge, MA: The MIT Press.

Zettersten, M., Schonberg, C. & Lupyan, G. (2020). What does a radical exemplar view not predict? A commentary on Ambridge (2020). *First Language* **40**(5/6), 636–639.

Appendix

TABLE 3. *Overview of traceback studies*

Study	Corpus / corpora	Language	Subcorpora	Operations	Semantic slot categories	Thresholds	Identification method	Overall traceback success
Lieven et al., 2003	Annie (2;0–2;1.11) 5 hours per week for six weeks	English	**test corpus:** last session of recording (all utterances) **main corpus:** remaining data incl. caregivers' input	substitute add-on drop insert rearrange	none (first analysis) OBJECT/ PERSON and LOCATION (second analysis)	1	semi-automatic	100% (N = 109) 79% of utterances used before, 21% novel utterances for novel utterances: 74% can be derived by 1 operation, 26% by 2+ operations
Dąbrowska & Lieven, 2005	Annie, Brian (2;0 and 3;0) 5 hours per week for six weeks	English	**test corpus:** last session of recording (only *wh*-questions) **main corpus:** remaining data, analysis 1: incl. caregivers' input; analysis 2: excluding caregivers' input	juxtaposition super-imposition	THING UTTERANCE PROCESS PROPERTY LOCATION DIRECTION	1 (first analysis) / 2 (second analysis)	not specified	traceback to all utterances incl. caregivers' input and frequency threshold of 2: Annie 2;0: 91% (N = 40) Brian 2;0: 91% (N = 10) Annie 3;0: 88% (N = 68) Brian 3;0: 89% (N = 39) traceback to child utterances only and frequency threshold of 2:

TABLE 3. *Continued*

Study	Corpus / corpora	Language	Subcorpora	Operations	Semantic slot categories	Thresholds	Identification method	Overall traceback success
								Annie 2;0: 61% (N = 27) Brian 2;0: 82% (N = 9) Annie 3;0: 60% (N = 46) Brian 3;0: 61% (N = 27
Lieven et al., 2009	Annie, Brian, Eleanor, Fraser, (2;0; Brian: 2;0–3;0) 5 hours per week for six weeks (Brian: for 1 year)	English	**test corpus:** last two hours (all utterances) **main corpus:** remaining data (excluding caregivers' input)	add substitute	REFERENT PROCESS ATTRIBUTE LOCATION DIRECTION UTTERANCE	1 (not explicitly specified)	semi-automatic	only graphic summary (Lieven et al., 2009, p. 493), no exact numbers provided
Vogt & Lieven, 2010	Annie, Brian, Eleanor, Fraser (2;0) 5 hours per week for six weeks	English	**test corpus:** last two hours (all utterances) **main corpus:** remaining data (including caregivers' input)	add substitute superimpose	REFERENT PROCESS ATTRIBUTE LOCATION DIRECTION POSSESSOR UTTERANCE	2	semi-automatic	only graphic summary (Vogt & Lieven, 2010, p. 26), no exact numbers provided
Kol et al., 2014	Adam (2;3–3;0), Eve (1;6–2;3), Nina (1;11-2;5), Sarah (2;3–2;7),	English	**test corpus:** last 10% of the child utterances	superimpose juxtapose	THING UTTERANCE PROCESS PROPERTY	2	automatic	(only results for traceback to all remaining utterances incl.

TABLE 3. *Continued*

Study	Corpus / corpora	Language	Subcorpora	Operations	Semantic slot categories	Thresholds	Identification method	Overall traceback success
	Thomas (3;0–3;2 and 3;3–4;0)		[analysis 2: with reversed word order; analysis 3: with randomized word order] **main corpus:** a) remaining utterances (incl. caregivers' input), b) remaining utterances (excluding caregivers' input), c) only child-directed speech utterances		LOCATION DIRECTION			caregivers' input) Adam: • normal: 85% • reversed: 68% • random: 65% Eve: • normal: 69% • reversed: 68% • random: 65% Sarah: • normal: 89% • reversed: 89% • random: 86% Nina: • normal: 88% • reversed: 85% • random: 87% Thomas-A (3;0–3;2) • normal: 69% • reversed: 57% • random: 59% Thomas-B (3;3–4;0) • normal: 60% • reversed: 49% • random: 50%
Miorelli, 2017	Roberto (2;1,13–2;2,26) 5 hours per week for six weeks	Italian	**test corpus:** last two hours (all utterances) **main corpus:** remaining data (including caregivers' input)	juxtaposition super-imposition	THING PROCESS PLACE SENTENCE QUALITY others	2	not specified	82% (N = 768)

TABLE 3. *Continued*

Study	Corpus / corpora	Language	Subcorpora	Operations	Semantic slot categories	Thresholds	Identification method	Overall traceback success
Koch, 2019	Marieke (2;2,22), Merit (2;0,21), Simon (2;04,23) 5 hours per week for seven weeks + part of the Leo corpus (2;03,13– 2;05,00)	German	**test corpus:** last two hours (all utterances) **main corpus:** remaining data (including caregivers' input)	add substitute superimpose	REFERENT PROCESS ATTRIBUTE LOCATION DIRECTION	2	semi-automatic	Leo: 86% (N = 97) Marieke: 88% (N = 154) Merit: 85.5% (N = 416) Simon: 95% (N = 95)

Note. Frequency threshold for establishing component units, as discussed in Section 2.2.3. Boldface indicates lexical overlap between two component units.