Individual differences in discourse priming

A traceback approach

Nikolas Koch, Antje Endesfelder Quick
and Stefan Hartmann
Ludwig-Maximilians-Universität München |
University of Leipzig | University of Düsseldorf

In this paper we use corpora of four monolingual German-speaking children at 2 years of age to analyze the effect of input on the activation of chunks and frame-and-slot patterns. For this purpose, we first investigate to what extent chunks and patterns can be traced back to the direct input compared to input which is not part of the immediate discourse situation. Second, we take mean length of utterance (MLU) into account to see how the level of proficiency influences the amount of priming in each child. Results indicate that children with a lower MLU rely more on priming than children who are more proficient. This conclusion is consistent with the usage-based assumption that children's linguistic development starts with a strongly item-based reproduction of input patterns that gradually gives rise to increasingly creative and productive uses of constructions.

Keywords: language acquisition, construction priming, traceback, usage-based model

1. Acquiring inventories: A usage-based perspective

Productivity is a crucial hallmark of language acquisition, which ultimately begs the question of how children acquire such productivity moving from single words to multiword utterances. Following a usage-based approach to language acquisition (Tomasello 2003), linguistic knowledge is acquired on the basis of the children's specific experiences with the world and the language(s) that surrounds them. Studies have shown that lexically specific units (*what's this?, let's go*) and frame-and-slot patterns ([*what's* X], [*let's* X]) feature heavily both in child-directed speech and in the early speech of children (Arnon and Christiansen 2017; Tomasello 2003). On their way to full mastery of their language, children's inventories are closely tied to what

https://doi.org/10.1075/bjl.00045.koc
Belgian Journal of Linguistics Volume 34 (2020), pp. 186–198. issn 0774-5141 | e-issn 1569-9676

they hear from their immediate caretakers (e.g. Cameron-Faulkner, Lieven, and Tomasello 2003; Kirjavainen and Theakston 2011; Theakston and Lieven 2008). Child-directed speech (CDS) is highly repetitive and also exhibits a high number of lexically restricted utterance-initial patterns, e.g. [*What did* X], which helps children acquire frame-and-slot patterns (Cameron-Faulkner, Lieven, and Tomasello 2003).

Immediate discourse effects also play an important role for production. For example, Kirjavainen and Theakston (2011) showed that infinitival *to* omission errors are subject to discourse priming of two related constructions. Rowland et al. (2012), investigating constructional alternations, showed a structural priming effect across different age groups, which was largest with the youngest children. This implies that younger children directly infer structural knowledge from the input they receive. It stands to reason that immediate structures in the input influence children's production of lexically specific units (*what's this?*) and frame-and-slot patterns ([*what's* X]). The younger children are, the more they rely on what the input can provide – the more proficient they get, the more they emancipate themselves from the input and start to use linguistic constructions productively (e.g. Rowland et al. 2012).

It is also clear that no child lives the same life, and differences in the in- and output are expected. They are even predicted in usage-based approaches as they assume that children build up linguistic knowledge on the basis of their experience, and that any subsequent behavior is based on this knowledge (e.g. Dąbrowska 2012). Consider, for example, vocabulary development: early lexical development is slow for all children, but very young children differ enormously in how they acquire their words (Frank et al. 2017). This depends on many factors such as individual processing abilities, socio-economic background, but also on the amount and quality of their linguistic experience. Since vocabulary and grammatical development are interdependent, it is not surprising that these differences also project into grammatical competence in this age range, which entails that some children are more proficient than others.

Keeping input, priming and individual differences in mind, the present paper investigates to what extent patterns can be traced back to the immediate parental input compared to parental input which is not part of the immediate discourse situation. Secondly, we are interested in whether the level of proficiency, calculated in Mean Length of Utterance (MLU), has an influence on the amount of priming in each child.

2. Tracing back utterances

The traceback method established by Lieven and colleagues (Lieven et al. 2003; Lieven, Salomo, and Tomasello 2009) provides a possible operationalization for analyzing patterns in the in- and output. The gist of the traceback method is to identify patterns by tracing back utterances to potential precedents in a corpus of naturalistic child speech collected via recordings. The usual procedure is to divide a corpus into a *test corpus* (usually the last two recording sessions) and a *main corpus* (the remaining data). Utterances in the test corpus are traced back to the previous utterances in the main corpus, and if the traceback yields a complete match, a lexically fixed pattern is established (e.g. *what's this?*). If only a part of an utterance is traced back, a frame-and-slot pattern is established (e.g. *I want X*). In order to constrain which frame-and-slot patterns can be posited, a set of operations is defined that can be used to systematically "construct" the target utterances from building blocks (so-called component units) in the main corpus (see e.g. Dąbrowska and Lieven 2005). In particular, three operations are typically used:

- ADD: juxtaposition of component units, e.g. *What's this* + *Mommy* = *What's this mummy*.
- SUBSTITUTE: one unit is replaced by another unit. Let us assume that we trace back the target utterance *What's this* and we don't find a verbatim match in the main corpus. However, we do find *What's that*. In this case, the operation SUBSTITUTE is used (substitute *that* by *this*) to derive the target utterance, and a frame-and-slot pattern [*What's* X?] is posited.
- SUPERIMPOSE is similar to SUBSTITUTE, but in this case, there is a partial overlap between the lexically fixed part of the frame-and-slot pattern and the component unit. For example, Vogt and Lieven (2010) trace back the target utterance *I can't open it*. In the main corpus, they find the fixed string *can't open it* but not *I can't open it*. However, they do find instances of [*I can't* X *it*] with other slot fillers, e.g. *I can't sing it*. Thus, they assume that the child has the frame-and-slot pattern [*I can't* X it] as well as the fixed string *can't open it* available.

Using this data-driven approach, the traceback method has shown that the vast majority of children's utterances can be accounted for on the basis of a limited set of patterns. In Section 3, we draw on a modified version of the traceback method by using the child's utterances from two recording sessions as target utterances and tracing them to their caregivers' utterances (a) in the same two recording sessions and (b) in two other recording sessions. This can give clues to the role of discourse priming, as well as effects of the situational contexts more generally, in children's choice of words and constructions.

In doing so, we want to fill a gap in previous work on first language acquisition using the traceback method. One consideration that has, to our knowledge,

not been sufficiently taken into account yet is that the fixed chunks and frame-and-slot patterns that are observed in the data may partly depend on contextual factors. In the present paper, we are explicitly interested in the role of the direct discourse situation. We are further interested in individual differences. Assuming that children with a lower MLU are more reliant on what the discourse can provide, we expect that children with a higher MLU will be less dependent on the immediate discourse.

3. Case study

3.1 Data

For the present study we used four corpora of German-speaking children aged between 2;0 and 2;6 years. Recordings for each corpus were made over a period of seven weeks with five recordings each week during typical play interactions each lasting one hour. As such the recordings capture 5 to 10% of daily speech (see e.g. Lieven, Salomo, and Tomasello 2009, 493). Table A in the Appendix gives an overview of the corpora.[1]

3.2 Method

As mentioned above, each corpus was divided into a test and a main corpus. Taken together, the four test corpora consist of 874 target utterances. Following Koch (2019), only utterances with three or more words were used as target utterances. In keeping with the operationalization outlined in Section 2, these target utterances were then traced back, but in contrast to previous traceback studies, they were not traced back to all previous recording sessions except for the last two, and not to the child's own utterances, but rather to (a) the input utterances in the last two recording sessions (i.e. the same ones from which the target utterances were taken), (b) the input utterances in the two previous recording sessions (functioning as a baseline to assess the influence of recency).[2]

---

1. For further details on the test persons, data collection, transcription and data preparation see Koch (2019, 157–170).

2. For details on the operationalization, see Koch (2019). Following Vogt and Lieven (2010), we use the operations SUPERIMPOSE, SUBSTITUTE, and ADD for identifying frame-and-slot patterns. As we are interested in direct priming patterns, we accepted any string or frame-and-slot pattern as a component unit that occurs at least once in the main corpus (while most other traceback studies work with a threshold of two occurrences).

In the following utterances provided by Leo and his mother, we exemplify how contextual factors influence the production of utterances. In Example (1a), we see the priming of a complete chunk (*guck mal, hier*), whereas Example (1b) shows the priming of a frame-and-slot pattern (*noch ein X*).

(1) a. MOT: ***guck mal*** , ***hier*** .
look PART here
'look, there'
CHI: ***guck mal hier***
b. MOT: *dann hole ich dir* ***noch ein*** *Bärchen.*
then fetch I you$_{DAT}$ another bear$_{DIM}$
'Then I'll get you another little bear'
CHI: ***noch ein*** *Hänger.*
another trailer
'another trailer'

In order to assess the child's current language development, we also calculated the mean length of utterance (MLU) for the last recording days under investigation in each child (see Table A in the Appendix). Following Behrens (2006, 11), MLU was measured in words (rather than morphemes).

3.3 Results

As mentioned above, our goal in using the traceback method in the way described above is to explore to what extent the immediate input correlates with the children's use of patterns. In addition, we discuss whether the degree of these correlations is related to children's individual language development by taking their MLU into account, which is often seen as an important measure of linguistic development.

Our first analysis, i.e. tracing back the children's target utterances to the caregivers' utterances in the two penultimate recording sessions, functions as our 'baseline'. Since the penultimate sessions are not part of the immediate discourse situation, we expect to find a lower priming rate for these utterances. This allows us to get an approximation to what extent the target utterances are related to the input in general, regardless of the immediate discourse situation.

Figure 1 shows the results of both traceback studies. The bars show the percentage of utterances successfully derived with the respective number of operations: For example, in the case of Leo, 4 of his 113 target utterancesare successfully traced back to the input data from the two previous recording sessions with 0 operations (upper left panel in Figure 1), i.e. there are verbatim matches in the input data used as main corpus. 12 utterances are derived using exactly one of the three operations mentioned in Section 2 (see Table A in the Appendix for the

number of target utterances for the other children). Taken together, this amounts to 10.6% of target utterances that were successfully traced using one operation or less. This is shown by the blue dots: The dots (and the line connecting them) show the cumulative traceback success.

The utterances could only be successfully traced back to the input data from the previous recording sessions in 18% to 49% of all cases (Figure 1, left panel). In contrast, when we trace the target utterances to the immediate discourse, these values increase to 44%–73% (Figure 1, right panel), which indicates that priming plays a substantial role. Mixed-effects binomial logistic regression models fit to the data for each child show that the traceback success significantly decreases in the case of Leo, Marieke, and Simon (see coefficients in Table 1). In the models, we used outcome (traceback success vs. fail) as the response variable and type (different vs. same session) as the only predictor variable. We also included the target utterance as a random effect (varying intercepts).

Only in the case of Marieke, the number of successful tracebacks remains nearly unchanged. A potential explanation for this observation is that she uses a large number of question constructions (see Koch 2019), which are highly formulaic. Thus, it makes sense that many of her target utterances can be as successfully traced to input from the same recording session as to input from any other session.

Table 1. Coefficients of the four regression models

	Estimate	Std. Error	Z value	Pr(>|z|)
Leo				
(Intercept)	0.56	0.28	1.98	0.048*
type:different_session	−2.1	0.47	−4.6	0.0000048***
Marieke				
(Intercept)	0.017	0.21	0.08	0.94
type:different_session	−0.099	0.25	−0.39	0.69
Merit				
(Intercept)	−0.38	0.13	−2.9	0.004**
type:different_session	−1.7	0.21	−8.2	3.1E-16***
Simon				
(Intercept)	1.64	0.46	3.59	0.00033***
type:different_session	−2	0.51	−3.9	0.000099***

Relating the children's language development to their rates of priming, we find individual differences: The lower the MLU is (see Table A in the Appendix for the MLU of all children), the more the children rely on what the discourse can pro-

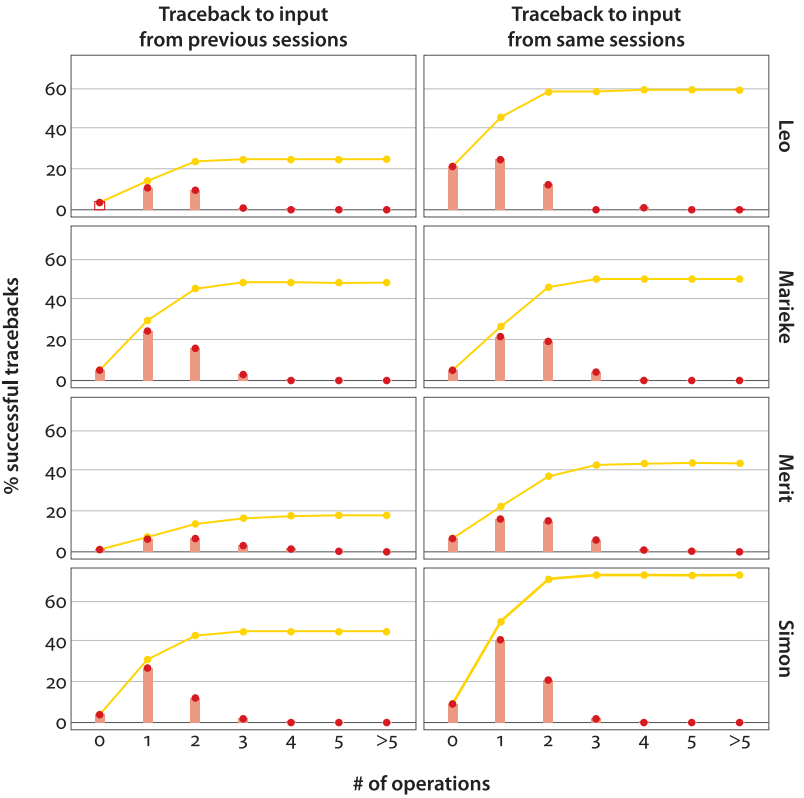

Figure 1. Proportion of target utterances successfully traced to (a) two previous recording sessions (left), (b) the last two recording sessions

vide. For example, the MLU of Simon's test corpus (1.8) is substantially lower than Merit's (4.8), and in turn, he relies more on priming (73%) than Merit (44%) does. This is consistent with the idea that children with lower proficiency use the discourse as a springboard to communication, whereas children with a higher proficiency can rely more on their own resources to structure their utterances.

4. Discussion and conclusion

The aim of the current study was to investigate the role of discourse priming in children's use of constructions. We further related children's language development as witnessed by their MLU to their reliance on discourse items. The results indicate that the structure of the preceding discourse plays a substantial role in children's construction of utterances in that they often used structures which were immediately occurring in the previous discourse. We further found that children's

use of discourse priming links up nicely with their MLUs: lower MLUs showed an increase of priming whereas children with a higher MLU did not rely on discourse that much anymore.

As a caveat, it should be pointed out that the traceback method that we used can only provide a rough proxy to the role of discourse priming as it only allows for comparing the degree of traceback success. But the traceback results help detect interesting patterns in the data which can serve as a starting point for subsequent, more detailed analyses. Also, investigating whether the differences in traceback success are mainly due to lexical (a lexical item is not part of the main corpus) or syntactic (syntagmatic differences) reasons would require a more in-depth investigation of the data in the future.

All in all, our results are in line with the key assumptions of constructionist approaches to language acquisition and consistent with previous findings on discourse priming (Pickering and Ferreira 2008). In research, priming has been identified as a pervasive feature. From a psycholinguistic perspective, priming is handy since it eases processing of structures which have been activated before (e.g. Schmid 2020). In language acquisition scenarios, priming helps in building up proficiency (Savage et al. 2006) – and this is what the results of our study also indicate: children are able to extract patterns from the input to use it in their own utterances. These input-output relations are predicted from a usage-based perspective, which assumes that children learn language primarily through experience (rather than relying on an innate Universal Grammar). As no child experiences the same input situations, a usage-based approach expects that each child extracts and builds up linguistic knowledge in at least slightly different ways, even though there are of course many commonality paths of acquisition as well. What usage-based approaches further predict is that children make use of patterns from the immediate discourse to differing degrees. In our study, we correlated utterance length as a measure of proficiency with the amount of priming as indicated by the traceback results. Children with lower proficiency, as witnessed by their lower MLU, seem to rely more on structures provided by the input in the immediate context. For example, Simons's MLU hovered around 1;8 and he relied on priming in 73% of the cases. (2) exemplifies Simon's reliance on fixed strings from the immediate discourse.

(2) MOT: *so,* ***da*** ***liegt*** *noch* *einer.*
so there lies another one
'So, there's another one'
MOT: ***da*** ***liegt*** *ja* *auch ein Eisbär.*
there lies PART also a polar-bear
'there's also a polar bear lying there'

CHI: ***da*** ***liegt*** *das.*
there lies this
'there it is lying'

Merit, on the other hand, is the most proficient child in our sample with an MLU of 4;8, and her traceback success is the lowest (44%) of all four children. In contrast to Simon, who increasingly takes up patterns of his caretakers from direct discourse and modifies them only slightly, Merit detaches herself more strongly from them. For example, the target utterance in (3a) has no direct precedent in the immediate context, and even the chunk *ich male* 'I paint' doesn't occur in the main corpus. Still, the utterance can be traced back to constructions used by Merit's mother such as those in (3b).

(3) a. *CHI: *ich male auch einen Wurm.*
I paint also a worm
'I also paint a worm.'
b. *MOT: *das Schneeflöckchen hat auch einen weinenden Mund?*
the snowflake$_{\text{DIM}}$ has also a crying mouth
'The snowflake also has a crying mouth?'
*MOT: *Maria hat auch einen dicken Bauch.*
Maria has also a fat belly
'Maria also has a big belly.'

As the words *ich, male*, and *Wurm* are also used in the main corpus, (3a) can be accounted for successfully by positing a pattern [X VERB *auch einen* Y] that is filled with the concrete lexical items, but in contrast to Simon's example in (2), abstract syntactic constructions play a much larger role here.

What these examples suggest is that discourse, especially at lower proficiency levels, functions as a gateway to communication, and it helps children to gradually increase their own proficiency. This is in line with the usage-based account of language acquisition. While strongly entrenched units can easily be retrieved from long-term memory (see e.g. Schmid 2020), less entrenched ones can be activated by discourse priming (and also become more entrenched in this process). Thus, discourse priming may act as a booster for the entrenchment and use of fixed chunks and frame-and-slot patterns. As children grow older, they start to use the patterns that they have acquired more productively without relying as much on patterns from the immediate context. Also, it has been shown that over the course of their linguistic development, children rely less on fixed chunks and more on frame-and-slot patterns (e.g. Bannard, Lieven, and Tomasello 2009).

As mentioned above, the traceback results provide support for these hypotheses, but of course they can only serve as a first step towards more detailed corpus-based analyses that could complement previous work in discourse priming in first

language acquisition, which has mostly used experimental methods or elicited data. Most importantly, we have argued that a corpus-based analysis of discourse priming can help us understand in more detail how children gradually build up the constructional network that constitutes their linguistic knowledge.

References

Arnon, Inbal, and Morten H. Christiansen. 2017. "More Than Words: The Role of Multiword Sequences in Language Learning and Use." *Topics in Cognitive Science* 9(3): 542–551. https://doi.org/10.1111/tops.12274

Bannard, Colin, Elena Lieven, and Michael Tomasello. 2009. "Modelling Children's Early Grammatical Knowledge." *Proceedings of the National Academy of Sciences* 106(41): 17284–17289. https://doi.org/10.1073/pnas.0905638106

Behrens, Heike. 2006. "The Input-Output Relationship in First Language Acquisition." *Language and Cognitive Processes* 21: 2–24. https://doi.org/10.1080/01690960400001721

Cameron-Faulkner, Thea, Elena Lieven, and Michael Tomasello. 2003. "A Construction Based Analysis of Child Directed Speech." *Cognitive Science* 27(6): 843–873. https://doi.org/10.1207/s15516709cog2706_2

Dąbrowska, Ewa. 2012. "Different Speakers, Different Grammars: Individual Differences in Native Language Attainment." *Linguistic Approaches to Bilingualism* 2 (3): 219–253. <https://www.jbe-platform.com/content/journals/10.1075/lab.2.3.01dab>. https://doi.org/10.1075/lab.2.3.01dab

Dąbrowska, Ewa, and Elena Lieven. 2005. "Towards a Lexically Specific Grammar of Children's Question Constructions." *Cognitive Linguistics* 16(3): 437–474. https://doi.org/10.1515/cogl.2005.16.3.437

Frank, Michael, Mika Braginsky, Daniel Yurovsky, and Virginia Marchman. 2017. "Wordbank: An Open Repository for Developmental Vocabulary Data." *Journal of Child Language* 44 (3): 677–694. < https://doi.org/10.1017/S0305000916000209>

Kirjavainen, Minna, and Anna Theakston. 2011. "Are Infinitival *to* Omission Errors Primed by Prior Discourse? The Case of WANT Constructions." *Cognitive Linguistics* 22(4): 629–657. https://doi.org/10.1515/cogl.2011.024

Koch, Nikolas. 2019. *Schemata im Erstspracherwerb: eine Traceback-Studie für das Deutsche*. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110623857

Lieven, Elena, Heike Behrens, Jennifer Speares, and Michael Tomasello. 2003. "Early Syntactic Creativity: A Usage-Based Approach." *Journal of Child Language* 30: 333–370. https://doi.org/10.1017/S0305000903005592

Lieven, Elena, Dorothé Salomo, and Michael Tomasello. 2009. "Two-Year-Old Children's Production of Multiword Utterances: A Usage-Based Analysis." *Cognitive Linguistics* 20: 481–507. https://doi.org/10.1515/COGL.2009.022

Pickering, Martin J., and Victor S. Ferreira. 2008. "Structural Priming: A Critical Review." *Psychological Bulletin* 134(3): 427–459. https://doi.org/10.1037/0033-2909.134.3.427

Rowland, Caroline F., Franklin Chang, Ben Ambridge, Julian M. Pine, and Elena V.M. Lieven. 2012. "The Development of Abstract Syntax: Evidence from Structural Priming and the Lexical Boost." *Cognition* 125(1): 49–63. https://doi.org/10.1016/j.cognition.2012.06.008

Savage, Ceri, Elena Lieven, Anna Theakston, and Michael Tomasello. 2006. "Structural Priming as Implicit Learning in Language Acquisition: The Persistence of Lexical and Structural Priming in 4-Year-Olds." *Language Learning and Development* 2(1): 27–49. https://doi.org/10.1207/s15473341lld0201_2

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198814771.001.0001

Theakston, Anna, and Elena Lieven. 2008. "The Influence of Discourse Context on Children's Provision of Auxiliary BE." *Journal of Child Language* 35: 129–158. https://doi.org/10.1017/S0305000907008306

Tomasello, Michael. 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Vogt, Paul, and Elena Lieven. 2010. "Verifying Theories of Language Acquisition Using Computer Models of Language Evolution." *Adaptive Behavior* 18(1): 21–35. https://doi.org/10.1177/1059712309350970

Appendix

Table A. Corpus overview

	Leo	Marieke	Merit	Simon
Recordings	2;03,13–2;05,00	2;02.22–2;04.10	2;00.21–2;02.07	2;04.23–2;06.18
Utterances main corpus (caretaker and child)	43544	43795	27089	28395
Utterances main corpus (child)	17813	17975	10204	8312
Utterances test corpus (caretaker and child)	2564	1539	1709	2491
Utterances test corpus (child)	1023	597	589	594
Number of target utterances (= child's multi-word units in the test corpus)	113	175	486	100
MLU test corpus (child)	1.617	2.379	4.847	1.822
Standard deviation MLU test corpus (child)	1.147	1.647	3.185	1.241

Authors' addresses

Nikolas Koch
Institute for German as a Foreign Language
Ludwig-Maximilians-Universität München
Ludwigstraße 27
80539 München
Germany
koch@daf.lmu.de

Antje Endesfelder Quick
Institute of British Studies
University of Leipzig
Beethovenstrasse 15
04107 Leipzig
Germany
antje.quick@uni-leipzig.de

Stefan Hartmann
Faculty of Arts and Humanities
University of Düsseldorf
Universitätsstraße 1
40225 Düsseldorf
Germany
hartmast@hhu.de