

Early Marking of Oblique Noun Phrases in German and English *

Sebastian Schuster

December 15, 2016

Abstract In this paper, I re-investigate the phenomenon that some English-speaking children use the preposition *from* to mark standards of comparison, demoted agents, and possessors, and I evaluate whether there is cross-linguistic evidence for children making use of a conceptual category of SOURCE (Clark & Carpenter 1989a) when formulating their early utterances. By comparing the early marking of standards of comparison in German and English through a corpus study and an elicitation task, I find that German-speaking children do not use *von* ('from') to mark standards of comparison, whereas some English-speaking children mark standards of comparison with *from*. Based on this result, I argue that English-speaking children overextend the conventional construction *different from* rather than relying on a conceptual category of SOURCE. I further present potential alternative explanations for the non-conventional uses of *from* to mark demoted agents and possessors.

1 Introduction

Grammatical prepositions such as *from*, *by*, and *with* can be used to mark a large number of semantic roles and their interpretation is highly context-dependent. Therefore, it is not surprising that children's early utterances sometimes contain non-conventional uses of these prepositions. What is more surprising, however, is that some of these non-conventional uses are very consistent. This is in particular true for the preposition *from*. Clark & Carpenter (1989a) noted that many children acquiring English use *from* to mark demoted agents (1), possessors (2), and standards of comparison (3):

- (1) a. I took my temperature **from** the doctor. (Julia, 2;2, talking about a visit to the doctor)
- b. I was caught **from** you before. (Jessica, 5;11, during a chase-game)

* I would like to thank my QP committee members, Beth Levin and Chris Manning, and in particular my QP chair, Eve Clark, for their advice and valuable discussions. I would also like to thank Meghan Sumner for her advice on developing the experiment, and Ed King, Masoud Jasbi, and Simon Todd for their feedback and suggestions at various stages of this project.

- (2) a. You can be a mom **from** two babies. (Damon, 3;7.5, assigning roles in a game)
- b. That's a finger **from** him (Shem, 3;7.5)
- (3) a. See? This ear is longer **from** the other ear. (Abe, 3;1.15, talking about a toy rabbit)
- b. Herb's the tallest **from** me and you're the tallest **from** me. (Damon, 3;7.1, comparing his height to that of his parents)

Adults would have used other prepositions in all of these sentences and [Clark & Carpenter \(1989a\)](#) therefore argued that these uses of *from* cannot be explained by the adult input. They argued that children subsume demoted agents, possessors, and standards of comparison as well as locative sources, temporal sources, and causes, under a general category of SOURCE, and therefore use the already acquired *from* to mark all of these types of arguments. As some of these arguments are conventionally not marked with *from* in English, children produce non-conventional utterances before acquiring the conventional preposition for each construction. The reasoning here is based on two observations. First, many languages extend spatial terms to non-spatial concepts and second, many languages express at least some of these arguments with the marker of locative sources. For example, within [Palancar's \(2002\)](#) sample of 98 languages, 30 languages use the marker of locative sources to mark demoted agents, and languages such as Uzbek and Estonian use the marker of locative sources for standards of comparison ([Stassen 1985](#)). The fact that many unrelated languages appear to conflate these types of arguments is taken as evidence for a conceptual category of SOURCE.

To the best of my knowledge, [Johnson \(1996\)](#) is the only other work that has investigated these non-conventional uses of *from*. Based on the assumption that children do not receive any form of negative feedback, he argued that children initially assign a very constrained meaning to *from*. Otherwise, if children assigned a more general meaning such as SOURCE to *from*, then they would subsequently have to refine their initial hypothesis, which can only happen when they receive negative feedback. He therefore suggested that children first assume that *from* marks only arguments which can be simultaneously viewed as locative sources, agents, and causes, and that they later learn that *from* can be used more generally, e.g., to mark locative sources independent of whether they can also be viewed as causes. However, one of his most important assumptions, the absence of negative feedback, has been disproved since then (see, e.g., [Chouinard & Clark \(2003\)](#), [Hiller & Fernández \(2016\)](#)) and therefore I will not discuss his account further in this paper.

There is some work on the acquisition of grammatical prepositions in English: Tomasello (1987) investigated his daughter’s first uses of prepositions and concluded that his daughter started using grammatical prepositions after she had begun to use spatial prepositions; and moreover, her first uses of grammatical prepositions were in combination with at least one other word, whereas her first uses of spatial prepositions came in the form of “holophrases”, i.e., the entire utterance consisted only of a preposition, as in *up* or *out*. However, these holophrases were typically used to describe or request an action such as picking something up, and therefore his daughter’s first uses of locative prepositions should probably be considered as uses of verbal particles. For this reason, it is not entirely clear whether there is a difference between the early uses of grammatical and spatial prepositions.

McKercher (2001) investigated the early uses of *with*. He concluded that children most likely assign a very general meaning of “having” to *with* based on several comprehension and production experiments. This account is in many ways similar to the account by Clark & Carpenter (1989a) as both of them assume that children initially assign a single underspecified meaning to a grammatical preposition.

More recently, Kidd & Cameron-Faulkner (2008) investigated all utterances containing *with* in a dense corpus from one English-speaking child. They did not find evidence for a very general meaning of *with* as proposed by McKercher (2001). Instead, they concluded that the child they investigated most likely first assigned a very constrained meaning of spatial proximity to *with* and then quickly progressed to a dynamic, context-dependent meaning that let him differentiate the various interpretations of *with* by age 3.

There is also more recent work on whether children possess abstract linguistic categories. For example, Tomasello (2000) drew upon previous results of diary studies and experiments and observed that up to age 3, children use verbs only with a very limited set of arguments that they have heard from adults. Based on this observation, he argued that young children lack abstract syntactic categories such as subject and object, as well as semantic categories such as agent and patient, because if children possessed these categories, we would expect them to be able to replace an argument with another argument of the same category and we would therefore observe many more different verb-argument combinations.

None of these more recent results necessarily contradict any of the assumptions made by the SOURCE account. First, neither work investigated the uses of *from* and second, even if these results apply to the acquisition of *from*, they mainly concern children who are younger than 3 and most children who produced non-conventional utterances with *from* did so after age 3. Nevertheless,

both raise the question of whether children make use of abstract conceptual categories in formulating their utterances.

In this paper, I therefore revisit the phenomenon of children’s overextensions of the preposition *from* and ask whether there is cross-linguistic evidence for children relying on a conceptual category of SOURCE to formulate their early utterances or whether it is more likely that children overextend specific constructions. For this purpose, I compare early uses of *from* by English-speaking children to early uses of *von* (‘from’) by German-speaking children. As I discuss in more detail in the following section, German appears to be particularly well-suited as a language for such a comparison as it is similar to English but it differs in how SOURCE-type arguments are marked.

This paper is structured as follows. I first compare how SOURCE-type arguments are conventionally marked in German and English (Section 2). I then present a corpus study as well as an elicitation experiment on the early marking of standards of comparison in the two languages (Section 3 and 4). Finally, I present potential alternative explanations for the non-conventional marking of demoted agents and possessors (Section 5).

2 Source Arguments in German and English

German-speaking and English-speaking adults mark SOURCE-type arguments in different ways. In this section, I discuss the conventional markers for each type of argument in turn.

2.1 Locative Sources

Locative sources in German are conventionally introduced by the prepositions *von* (‘from’) and *aus* (‘out of’). Both prepositions can be used to mark the source of a motion event (4-5) as well as where an object comes from (6-7)¹.

- (4) Der Apfel fiel **vom** Baum.
The apple fell from-the.SG.MASC.DAT tree
‘The apple fell from the tree.’
- (5) Sie geht **aus** dem Haus.
She walks out-of the house
‘She is walking out of the house.’
- (6) Der Apfel ist **vom** Bauernhof.
The apple is from-the.SG.MASC.DAT farm

¹ *Vom* in (4) and (6) is a contraction of *von* and *dem*, the singular, masculine determiner in dative case.

‘The apple comes from the/a farm.’

- (7) das Keks **aus** der Dose
the cookie out-of the jar
‘the cookie from the jar’

The difference between *von* and *aus* is that *von* typically marks a punctual source whereas *aus* is used when the source is a container or a delimited area such as a town or a country and the object used to be in the container or within the limits of that area. While this generalization holds for most conventional uses of these two prepositions, there are some exceptions to this rule. Most notably, whenever a source and a goal of a motion event are expressed in the same clause, *von* tends to be used.

In English, adults also use *from* and *out of* to mark locative sources. *Out of* is also used when the object is or was contained by the locative source. However, one contrast to German is that *out of* appears to be limited to motion events and cannot be used as a modifier of a noun phrase as in (8).

- (8) the cookie ***out of/from** the jar.

In addition, English has a third preposition to mark locative sources, namely *off (of)*, which can be used instead of *from* to highlight the fact that the source is a surface.

- (9) The workers jumped **off (of)** the back of the truck.

2.2 Temporal Sources

Starting points in time and temporal sources are also conventionally introduced by *von* in German and *from* in English. In German, this construction can be used to indicate when an object was made/acquired/received as in (10), to mark the starting point of a time span (11), and in combination with *an* (‘on’) to mark a starting point in time (12).

- (10) die Zeitung **von** gestern
the newspaper from yesterday
‘yesterday’s newspaper’
- (11) **von** Montag bis Donnerstag
from Monday until Thursday
‘from Monday through Thursday’
- (12) **von** morgen an
from tomorrow on
‘from tomorrow on/starting tomorrow’

German speakers also use the preposition *ab* (‘starting’, ‘from ... on’) to mark starting points in time. *Ab* can be used instead of *von ... an* as in (13), and it is frequently used in the phrase *ab X Jahren* (‘from X years on’) in board games and movie ratings to indicate a recommended age range (14).

- (13) **ab** nächster Woche
starting next week
'starting next week'
- (14) **ab** 3 Jahren
starting 3 years
'age 3 and older'

The complements of *von* can be temporal expressions as in (10-12) as well as prior events (15).

- (15) der Kuchen **von** meiner Geburtstagsfeier
 the cake from my birthday-party
 ‘the cake from my birthday party’

English speakers use *from* in a temporal sense primarily for starting points of time spans (16) and starting points in time (17). Further, *from* can be used in combination with events that serve as starting points (18) or the temporal reference point of an event (19).

- (16) She works **from** Monday through Friday.
- (17) **from** now until four o'clock (Clark & Carpenter 1989a)
- (18) **From** World War II on, their fortunes failed.
(Clark & Carpenter 1989a)
- (19) the cake **from** my birthday party

For certain common temporal expressions such as *yesterday* or *last week*, English speakers conventionally use a possessive construction instead of a prepositional phrase to mark the temporal origin (20-21).

- (20) yesterday's bread
(21) last week's news

Lastly, English speakers also use the participial preposition *starting* to mark a starting point in time.

- (22) **Starting** tomorrow, I will exercise more.

2.3 Demoted Agents

German speakers conventionally use *von* to mark demoted agents in passive constructions.

- (23) Das Buch wurde **von** Stefanie Sargnagel geschrieben.
The book became from Stefanie Sargnagel written
'The book was written by Stefanie Sargnagel.'

In English, on the other hand, *by* is used to introduce demoted agents. However, there are some exceptions to this rule. First, in the construction *get ... from*, *from* can be used to mark a locative source, which can also be viewed as an agent.

- (24) She got a book **from** John.
(25) He got a hug **from** Jill.
(26) Sue got a call **from** Mel.

Further, Clark & Carpenter (1989a) noted that in some dialects of English, *from* can also be used to mark natural forces.

- (27) The ice cream melted **from** the sun. Clark & Carpenter (1989a)
(28) The branch **broke** from the wind. Clark & Carpenter (1989a)

2.4 Causes

German speakers can use three constructions to express a causal relation. First, they can use the complementizer *weil* ('because'), which always requires a tensed clause as its complement (29). Second, they can use *von* ('from') with a nominal complement to express a non-agentive, direct² causal relation (30) (Maienborn & Herdtfelder 2015). Third, they can use *wegen* ('because of') with a nominal complement to express indirect causal relations or reasons (31).

- (29) Sie war müde **weil** sie lange gearbeitet hatte.
She was tired because she long worked had
'She was tired because she had worked long hours.'
- (30) Sie war müde **von** der Arbeit.
She was tired from work.
'She was tired from work.'

² See, for example, Wolff (2003) for a discussion of the differences between direct and indirect causes.

- (31) Sie fuhren **wegen** der Hitze an den See.
 They drove because-of the heat to the lake
 ‘They drove to the lake because of the heat.’

Apart from these three constructions, there also exist several idiomatic expressions that can be used to express a causal relation. For example, *sterben* (‘to die’) can be combined with a prepositional phrase headed by *an* (‘on’) to express the cause of death.

- (32) Er starb **an** einem Herzinfarkt.
 He died on a heart-attack.
 ‘He died from/of a heart attack.’

The construction *Das kommt von ...* (‘This/that comes from ...’), which occurs particularly often in child-directed speech, is another way of expressing cause. This construction is often used to explain a cause or reason to a child.

- (33) Das kommt **vom** Rumschleudern [...].
 That comes from-the.SG.MASC.DAT hurtling-around
 ‘that comes from hurtling around.’ (Mother to Leo, 2;3.20)

The means for expressing cause in English are very similar to the ones in German. *Because* is also used to express a causal relationship between two clauses (34), and *from* and *because of* are used with a nominal complement (35-36). However, causes of death are conventionally expressed with *from* or *of* in English (37).

- (34) She was tired **because** she had worked long hours.
 (35) She was tired **from** work.
 (36) They drove to the lake **because of** the heat.
 (37) He died **from/of** a heart attack.

2.5 Possessors and Partitives

German offers several means of expressing possession. The prescriptively correct means are the genitive case (38), or, for named animate possessors, the clitic *-s* (39). However, in spoken language, two other options for expressing possession are more common: marking possessors with *von* (40) and including a possessive pronoun between the possessor and the possessum (41).

- (38) das Fahrrad **des** Historikers
 the bicycle the.SG.MASC.GEN historian.SG.MASC.GEN
 ‘the historian’s bicycle’

- (39) Ronalds Fahrrad
 Ronald's bicycle
 ‘Ronald's bicycle’
- (40) das Fahrrad **von** Ronald
 the bicycle from Ronald
 ‘Ronald's bicycle’
- (41) Ronald **sein** Fahrrad
 Ronald his bicycle
 ‘Ronald's bicycle’

Partitive constructions are expressed using either the genitive case (42) or *von* (43). This also holds for arguments of relational nouns (Barker 1995) as illustrated in (44) and (45).

- (42) einer **der** **Studierenden**
 one the.PL.MASC.GEN students.PL.MASC.GEN
 ‘one of the students’
- (43) einer **von** den Studierenden
 one from the students
 ‘one of the students’
- (44) die Zerstörung **der** **Stadt**
 the destruction the.SG.FEM.GEN city.SG.FEM.GEN
 ‘the destruction of the city’
- (45) das Foto **von** Julia
 the photo from Julia
 ‘the photo of/from Julia’

Note that the use of *von* to mark locative sources, agents, possessors, and arguments of relational nouns can lead to ambiguities. The German phrase in (45), for example, has four readings. Depending on the context, (45) can express that the photo depicts Julia (argument of the relational noun), that Julia took the photo (agent), that Julia owns the photo (possessor), or that the speaker got the photo from Julia (locative source).

Possessors in English are either marked with the clitic *'s* or the preposition *of*. The former is used with animate possessors and the latter with inanimate ones. Partitive constructions as well as the arguments of relational nouns are marked with *of*.

2.6 Standards of Comparison

In German, standards of comparison are conventionally marked with *als* (‘than’, ‘as’, ‘when’).

- (46) Markus ist größer **als** Arthur.
Markus is bigger than Arthur
‘Markus is taller than Arthur.’

In the case of equatives, *so ... wie* (‘so ... as’) is used to mark the standard.

- (47) Das Haus ist **so** groß **wie** ein Schloss.
The house is so big as a castle
‘The house is as big as a castle.’

However, speakers of some German dialects also mark standards of comparison with *wie* (but without *so* before the adjective) in comparisons of inequality³.

- (48) Sein Bruder ist kleiner **wie** er.
His brother is smaller as he
‘His brother is smaller than him.’

Apart from *als* and *wie*, many adults also use the archaic preposition *denn* in the fixed construction *X denn je* (‘X than ever’), e.g., *mehr denn je* (‘more than ever’).

In English, standards of comparison are marked with *than* as in the translation of (46). One exception to this rule is the adjective *different*, which is most often used with *from*, sometimes with *than*, and occasionally with *to*⁴. Equatives are marked with *as ... as* (49).

- (49) The house is **as** big **as** a castle.

2.7 Summary

Table 1 summarizes which prepositions and other case-marking mechanisms are conventionally used to mark SOURCE arguments. As this table shows, standard of comparison is the only SOURCE argument that is never marked with *von* in German. In English, demoted agents, possessors and partitives are never marked with *from*.

³ Historically, *als* used to be the equative marker and *denn* the comparative marker (Thurmair 2001), so the use of *wie* as a comparative marker could be the beginning of a similar shift in meaning.

⁴ Among the 20,449 uses of *different* with a standard of comparison in the Corpus of Contemporary American English (COCA), 75% mark the standard with *from*, 22% with *than*, and 2% with *to*.

	German	English
Locative sources	<i>von, aus</i>	<i>from, out of, off of</i>
Temporal sources	<i>von, ab</i>	<i>from, 's, starting</i>
Demoted agents	<i>von</i>	<i>by, (from)</i>
Natural forces	<i>von</i>	<i>by, from</i>
Causes	<i>von, weil, wegen, (an)</i>	<i>from, because, because of, (of)</i>
Possessors & partitives	genitive, -s, <i>von</i> , possessive pronoun	<i>'s, of</i>
Standards of comparison	<i>als, so ... wie, (denn), (wie)</i>	<i>than, as ... as, (from)</i>

Table 1 Summary of the conventional means for marking SOURCE arguments in German and English. Parentheses indicate non-standard uses or that the preposition can only be used in special contexts.

3 Corpus Study

As discussed in the previous section, standard of comparison (COMPARISON) is the only type of argument never marked with *von* in German. At the same time, English speakers mark COMPARISON in some cases with *from*. For this reason, the means that German-speaking children use to express COMPARISON could provide additional evidence for a conceptual category of SOURCE or could suggest that the non-conventional uses are limited to English-speaking children and are therefore language-dependent.

We know from [Clark & Carpenter \(1989a\)](#) that English-speaking children occasionally use *from* to mark COMPARISON in a non-conventional manner. [Clark & Carpenter \(1989a\)](#) attributed these non-conventional uses to children perceiving standards of comparison as abstract locative sources and therefore marking them with *from*. This argument is based on [Jackendoff \(1983\)](#), who argued that comparison can be perceived as an abstract movement away from the standard of comparison⁵. Crucially, this explanation implies that children’s non-conventional uses are independent of the linguistic input. If this is the case,

⁵ Alternatively, one can also perceive the standard of comparison as the viewpoint of the comparison, e.g., *an elephant is taller than a mouse* can be perceived as an elephant being taller from the viewpoint of a mouse. This is also reflected in the etymology of *than* and *denn*, which both go back to the gothic *þana*, *þan*, which in return can be traced back to the ablative of the demonstrative root *þa*- (von [Stuckrad 1957](#), [Thurmair 2001](#)).

Corpus	Age	Child utterances	<i>von</i>	<i>vom</i>	<i>als</i>
Leo	1;11.12 - 4;11.05	190,339	1,152	111	331
Cosima	0;00.13 - 7;02.22	26,138	148	29	60
Pauline	0;00.12 - 7;11.03	25,682	253	37	61
Sebastian	0;00.17 - 7;05.11	24,425	221	64	65
Caroline	0;10.01 - 4;03.18	24,982	90	26	22
Simone	1;09.11 - 4;00.06	35,107	97	16	16

Table 2 Summary statistics of the investigated German corpora. *von*, *vom*, and *als* denote the number of child utterances containing the relevant preposition.

some German-speaking children should also mark COMPARISON with *von* despite the fact that adults never use such a construction. (SOURCE HYPOTHESIS)

On the other hand, if German-speaking children never use *von* to mark standards of comparison, then this suggests that the non-conventional uses in English depend on the linguistic input and that most likely some English-speaking children overextend the construction *different from* to other comparative adjectives⁶. (CONSTRUCTION HYPOTHESIS)

These competing hypotheses make different predictions. According to the SOURCE HYPOTHESIS, as many German-speaking children should mark standards of comparison with *von* as English-speaking children mark COMPARISON with *from* because their production of this construction should be independent of the adult input.

According to the CONSTRUCTION HYPOTHESIS, German-speaking children should mark standards of comparison with *von* less often than English-speaking children mark COMPARISON with *from* because the marking of COMPARISON with *von* in the German input is never found, whereas COMPARISON is occasionally marked with *from* in the English input.

In order to test these predictions, I investigated children’s uses of *von* and *from* as well as *als* and *than* in several longitudinal corpora for German and English in the CHILDES Archive (MacWhinney 2000).

3.1 Data

I used the following six German corpora for my study. Table 2 contains the summary statistics for each corpus.

⁶ This explanation was also suggested by Johnson (1996), but he did not present any evidence for this hypothesis.

Leo The first corpus is a dense longitudinal corpus, collected by Behrens (2006), in the late 1990s and early 2000s. The corpus contains the transcripts of recordings of a child interacting with his parents and occasionally a research assistant at home. Between age 1;11.12 and 2;11.29, five hour-long sessions were recorded per week and supplemented with daily parental diary data noting the 10 to 30 most innovative and complex utterances of the child. Between age, 3;0 and 3;11, five hour-long sessions were recorded in every 4th week. Most of the recordings were made while Leo was playing with one of his caretakers.

Cosima, Pauline, and Sebastian The Cosima, Pauline, and Sebastian corpora were collected by Rosemarie Rigol in the 1990s. All three children were recorded for 30 minutes approximately every two weeks until age 4 and approximately every 4 weeks after that⁷. During the recording sessions, the children mainly interacted with Ms. Rigol, the investigator, but a parent, a sibling, or a friend of the child was also often present.

Caroline The Caroline corpus was collected by Christiane von Stutterheim in the late 1980s and early 1990s. The corpus is dense for the age range between 1;9 and 2;10 with an average of 14 recordings per month but after age 2;10, there are only 6 more recordings in total up to age 4;3. The recordings were made at home while the child was interacting with her parents or playing by herself.

Simone The Simone corpus was collected by Max Miller (1979) (the father) in the early 1970s. The length of recordings varies throughout the course of the study. Miller (1979) recorded Simone every six weeks for at least six hours. In between these six-week intervals, he made three additional shorter recordings (about one hour long). The recordings were made at home while Simone was interacting with her father and sometimes with her mother.

I also investigated the use of *from* and *than* in the following six English corpora⁸. The summary statistics of the English corpora are presented in Table 3. Additionally, I also considered the diary data on Damon for my analyses. However, as his data are not available through the CHILDES archive, I relied on the numbers reported by Clark & Carpenter (1989a).

⁷ Pauline was recorded in less regular intervals but still at least once a month during almost the entire period.

⁸ These are the same corpora as used by Clark & Carpenter (1989a).

Corpus	Age	Child utterances	<i>from</i>	<i>than</i>
Adam	2;03.04 - 5;02.12	46,743	142	23
Sarah	2;03.05 - 5;01.06	38,096	62	10
Eve	1;06.00 - 2;03.00	12,113	19	1
Shem	2;02.16 - 3;02.02	18,166	60	0
Abe	2;04.24 - 5;00.11	22,932	119	68
Naomi	1;02.29 - 4;09.03	17,242	18	0
Damon	1;06.00 - 5;11.00	~15,000	198	-

Table 3 Summary statistics of the investigated English corpora. *from* and *than* denote the number of child utterances containing the relevant preposition. The numbers for Damon are from [Clark & Carpenter \(1989a\)](#) as his data are not available through CHILDES.

Adam, Sarah, and Eve The Adam, Sarah, and Eve corpora were recorded by [Brown \(1973\)](#) in the early 1960s. Sarah was recorded for one hour approximately every week, and Adam and Eve were recorded for one hour approximately every two weeks. In all the recordings, the child is either interacting with a parent or with a research assistant.

Shem The Shem corpus was collected by [Clark \(1978\)](#) in the 1970s. Shem was recorded approximately once a week for one hour. In most of the recordings, he is interacting with a research assistant and one of his parents at home.

Abe The Abe corpus (also known as Walt) was collected by [Kuczaj \(1977\)](#) (his father) in the early 1970s. Kuczaj made two 30 minute recordings per week from age 2;4 to age 4;1, and one 30 minute recording per week from then on. The corpus contains transcripts of Abe interacting with his father at home.

Naomi The Naomi corpus was collected by [Sachs \(1983\)](#) (her mother) in the early 1970s. The recording intervals of this corpus vary a lot. Before age 1;8, Sachs made only two recordings; between age 1;8 and 2;11, she made approximately one recording every two weeks with occasional short periods of daily recordings in between, and from age 2;11 to age 3;9, she made approximately one recording per month. Over the course of the last year, she recorded Naomi only three more times. In all recordings, the child is interacting with her mother and occasionally with her father at home.

Damon The diary data on Damon was collected by Eve Clark (his mother). The diary contains utterances by Damon, including very detailed linguistic and non-linguistic context information, from daily observations up to age 6.

In summary, the German and English corpora are easily comparable. All of them were collected at the child’s home and in all of them, the child is interacting either with a parent or a researcher. The German corpora contain on average slightly more utterances than the English ones, mainly due to the fact that most of the German corpora were recorded over a longer period of time. In terms of density, however, the German and English corpora are very similar with the exception of the dense German Leo corpus for which no comparable English corpus is available.

3.2 Statistical model

One of the challenges of working with child data is that one often has to draw conclusions from very small sample sizes. Even for high-resource languages such as English and German, the CHILDES archive only contains a handful of suitable longitudinal corpora. At the same time, the collection of elicited data in experiments is both time consuming and logistically challenging, which is why it is often also not possible to collect data from a large number of children. For this reason, it is desirable to give a statistical framework that allows one to incorporate evidence from multiple data sources. I therefore test my predictions using a Bayesian statistical framework.

One of the advantages of Bayesian statistics is that one can specify a prior probability for each hypothesis based on previous observations (see, for example, [Efron \(2005\)](#)). If no previous observations are available, one typically assumes that the prior probability of competing hypotheses is the same, i.e., for two hypotheses, $P(H_0) = P(H_1) = 0.5$. However, if previous observations are available, one can incorporate these as a prior into another hypothesis test. This can be best illustrated with an example. Assume that we want to investigate whether a certain construction is used by more children in group *A* than children in group *B* (the alternative hypothesis). For each group of children, we have five corpora available and 4 children in group *A* and 2 children in group *B* use the construction being investigated. Analyzing these data with equal priors suggests that the alternative hypothesis is only approximately twice as likely as the null hypothesis, which would not be significant. However, based on this analysis, our beliefs about the two hypotheses have shifted, and now $P(H_0) \approx \frac{1}{3}$ and $P(H_1) \approx \frac{2}{3}$. If we then collect data from five more children in each group through an elicitation task, and 3 children in group *A* and 1 child

in group B use the construction, then we get a significant result if (and only if) we incorporate the information from the corpus data as a prior. However, if the elicited data was unlikely under H_1 , then the posterior probability of H_1 would be lower and we would not get a significant result, as expected. This example shows that in a Bayesian statistical framework, one can draw conclusions from small samples by incorporating multiple sources of evidence even if some or all of the individual evidence would not allow one to draw these conclusions.

So far, I have not mentioned how the significance of a result is determined. Instead of a p -value, which is normally used in a frequentist framework, one computes the Bayes factor (BF):

$$BF = \frac{P(D|M_1) P(M_1)}{P(D|M_0) P(M_0)},$$

where D is the observed data and M_0 and M_1 are the models corresponding to H_0 and H_1 , respectively. The model for a hypothesis specifies how the data are generated according to the hypothesis. For example, a reasonable model corresponding to the hypothesis that a series of h heads and t tails was obtained with a fair coin would be that the number of heads h is drawn from a binomial distribution parameterized by $\theta = 0.5$ (the probability of heads) and $n = h + t$ (the number of trials). $P(D|M)$ is then the probability of generating D according to M , which can be either computed exactly or by using an approximate inference algorithm such as Markov chain Monte Carlo (MCMC). In order to compute the Bayes factor, one computes $P(D|M_1)$ and $P(D|M_0)$ and combines them with the ratio of the priors $\frac{P(M_1)}{P(M_0)}$. The Bayes factor will then be a value in the interval $(0; \infty)$. A Bayes factor between 1 and 3 is considered anecdotal evidence for the alternative hypothesis, a value between 3 and 10 is considered moderate evidence, and a value above 10 is considered strong evidence for the alternative hypothesis (Jeffreys 1961, Lee & Wagenmakers 2013). In practice, a Bayes factor above 3 is considered significant.

Returning to my research question, the models corresponding to the SOURCE HYPOTHESIS and the CONSTRUCTION HYPOTHESIS are shown in Figure 1. Intuitively, we can imagine the generative process of the data as following. Whenever a child is born, a nurse flips a coin with an unknown probability θ of landing on head and if the coin lands on head, then the child will use *von* or *from* to mark standards of comparison. In the model on the left, the nurse uses the same coin for all children, whereas in the model on the right, the nurse uses a coin with probability of heads of θ_1 for all German-speaking children and another coin with probability of heads of θ_2 for all English-speaking children. Given the number of German-speaking children who use *von* to mark COMPARISON (s_1), the number of English-speaking children who use *from* (s_2),

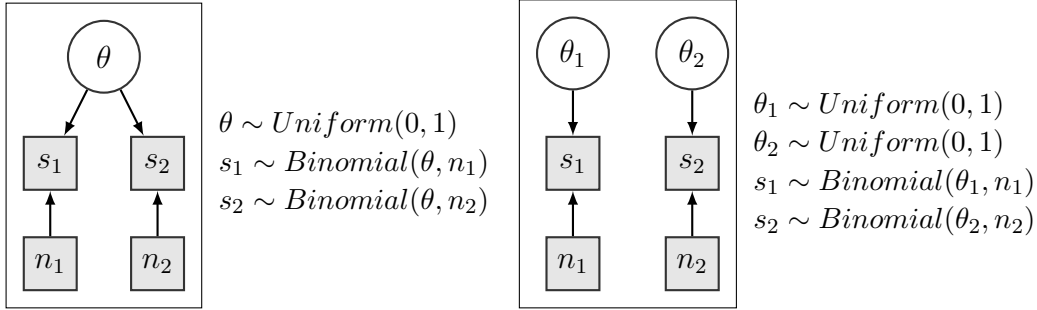


Figure 1 The generative models used for model comparison. In the left model, the latent variable θ is the same for both languages whereas in the right model, θ is different for each of the two languages.

and the total number of German-speaking and English-speaking children who use a comparative construction in the corpus (n_1 and n_2), we can then compute and compare the likelihood of the two models. If it is more likely that the nurse used the same coin, then the data suggest that there is no difference between the two language groups and this would be evidence for the SOURCE HYPOTHESIS. However, if it is more likely that the nurse used two different coins, then the data suggests that there is a difference between the two language groups and this would be evidence for the CONSTRUCTION HYPOTHESIS.

I implemented these two models in webppl (Goodman & Stuhlmüller 2014). I use MCMC as the inference algorithm and run it with 1,000,000 samples and a burn-in period of 50,000 samples.

3.3 Results

Table 4 shows the frequency of utterances with standards of comparison in the German and English corpora. I manually coded all utterances that contain *von/from* or *als/than* for whether they contain a comparative construction or not⁹. There was at least one use of a comparative construction with *als* in all German corpora but no German child produced a comparative construction with *von*. In English, on the other hand, only five children ever expressed COMPARISON and three of them used *from* to mark a standard of comparison. If we run the statistical model from the previous section with these numbers, we get a BF of 4.6, which suggests that significantly fewer German-speaking

⁹ In order to account for the larger age range of some of the German corpora, I only considered utterances until age 5;0.0.

	Leo	Cosima	Pauline	Sebastian	Caroline	Simone	
Age of first use	2;3	3;10	2;2	3;7	2;7	2;2	2;9
<i>von/vom</i>	0	0	0	0	0	0	0
<i>als</i>	58	1	4	5	11	12	91
<i>von/vom</i>	×	×	×	×	×	×	0
<i>als</i>	✓	✓	✓	✓	✓	✓	6

	Damon	Adam	Sarah	Eve	Shem	Abe	Naomi	
Age of first use	2;8*	4;2	2;10	2;1	-	3;3	-	3;0
<i>from</i>	18	0	2	0	0	9	0	29
<i>than</i>	-*	21	10	1	0	70	0	102
<i>from</i>	✓	×	✓	×	×	✓	×	3
<i>than</i>	-*	✓	✓	✓	×	✓	×	4

Table 4 Uses of standards of comparison in the German (top) and English (bottom) corpora. *Age of first use* denotes the age of the first use of a standard of comparison, the numbers in the second and third row show the absolute uses of the respective preposition to mark COMPARISON, and the last two rows indicate whether the child used the respective preposition to mark COMPARISON. The last column shows the average age and the row totals, respectively. *The data for Damon is from [Clark & Carpenter \(1989a\)](#), who only provide the number of utterances with COMPARISON marked with *from* along with several examples.

children use *von* to mark COMPARISON than English-speaking children use *from* for these constructions.

3.4 Discussion

The results of this corpus study suggest that only English-speaking children produce standards of comparison that are marked with *from*. This provides evidence for the CONSTRUCTION HYPOTHESIS as we would not expect a difference between the two languages if these non-conventional productions were independent of the adult input. Further evidence for the CONSTRUCTION HYPOTHESIS comes from the fact that Sarah and Abe used the conventional

comparative *different from* at least once¹⁰, which suggests that they picked up this construction from adult uses with *different*.

As shown in Table 4, Sarah used *from* only twice to mark a standard of comparison. Her first utterance contained the conventional construction *different from* (*dance a different way from us but very nice to see*, 3;10.16). Her second utterance contained *from* in combination with the comparative adjective *higher* (*now the rainbow is gettin(g) higher from the rain*, 5;0.30). However, with this example, it is unclear whether she wanted to express that the height of the rain is a standard of comparison or that the rain causes the rainbow to become higher. Abe, on the other hand, used this construction much more productively with adjectives such as *better*, *longer*, and *same*, and at some point even with the noun *winner* ([...] *I think we're gonna be a winner from Joey*, 3;8.28).

Abe shows a lot of consistency in his uses of the two prepositions. He only used *from* to mark COMPARISON until age 3;3 and from then on he exclusively used *than*, with the exception of the example with the noun phrase above. This is consistent with either hypothesis. According to the SOURCE HYPOTHESIS, he would first assume that standards of comparison are a source and then later learn to distinguish them from other arguments. According to the CONSTRUCTION HYPOTHESIS, one possible developmental path is that he starts with the assumption that *from* marks COMPARISON because of hearing *different from* and later learns that *from* can be used only with one particular adjective.

The adult input is also consistent with the CONSTRUCTION HYPOTHESIS. All corpora except the Naomi corpus contain at least one adult use of the phrase *different from*¹¹, and therefore all children who marked COMPARISON with *from* heard this construction at least once.

Additional evidence for children picking up non-default means of marking COMPARISON from the adult input comes from the German data. As I mentioned in Section 2, German-speaking adults sometimes mark COMPARISON with *wie*, which is prescriptively used only to mark equatives, as well as with *denn*, which is an archaic marker and only being used in a limited number of fixed constructions in modern German. These two prepositions also show up in children's early productions (50-51).

¹⁰ Clark & Carpenter (1989a) do not mention whether Damon also used this construction at one point.

¹¹ No input data are available for Damon but considering that *different from* is even more prevalent in British English and considering that his mother is a speaker of British English, it is very likely that he was exposed to this phrase.

- (50) aber des is(t) doch bißchen kleiner **denn** mir, so ein Loper. der
 but the is after-all a-bit smaller than me such a Loper. the
 is(t) kleiner **als** mir.
 is smaller than me.
 ‘But after all, this is a bit smaller than me, such a Loper¹². This is
 smaller than me.’ (Simone; 2;11.18)
- (51) größer **wie** die Giraffe.
 taller/bigger as the giraffe.
 ‘taller as the giraffe.’ (Simone; 3;01)

These examples suggest that German-speaking children also pick up non-default means and apply them more liberally than adults would do.

In summary, the corpus study provides evidence for the CONSTRUCTION HYPOTHESIS. However, considering the small sample sizes and the fact that children express standards of comparison infrequently, it could also be that German children use *von* to mark COMPARISON but none of these uses were captured on any of the recordings. In order to determine whether the results of the corpus study are consistent with additional data, I conducted an elicitation study, which I describe in the next section.

4 Elicitation Study

As I mentioned above, children only infrequently express standards of comparison in spontaneous speech and therefore it can be challenging to elicit uses of this construction. One proven method to elicit rare constructions is to ask children to imitate sentences that contain the construction of interest (Clark & Carpenter 1989b). When young children are asked to imitate a sentence, they frequently process the sentence through their own linguistic system and produce a version that is closer to what they would have said spontaneously than what they actually heard (Slobin & Welsh 1973). Therefore, if we ask children to imitate a sentence with a standard of comparison marked by *than*, we expect the child to change the preposition to *from* if he or she would also use *from* spontaneously for COMPARISON. At the same time, we can also ask children to imitate ungrammatical sentences with a COMPARISON marked by *from*. If they would spontaneously mark COMPARISON with *from*, then we expect them to imitate the sentence verbatim; otherwise we would expect them to substitute *than* for *from*.

While this method works well for young children up to around age 4;0, older children are a lot better at producing exact imitations (Clark & Carpenter

12 Presumably she is referring to a toy.

	<i>n</i>	female	male	age range	age mean	age SD (months)
< 4;6, ger.	7	3	4	3;04-4;03	3;10	3.8
> 4;6, ger.	8	6	2	4;09-5;11	5;04	5.2
< 4;6, eng.	9	3	6	2;08-4;02	3;07	6.1
> 4;6, eng.	11	4	7	4;06-6;01	5;04	7.3

Table 5 Sex and age statistics of subjects in each age and language group.

1989b) and therefore their imitations are not as much in line with what they would say spontaneously. However, older children’s meta-linguistic abilities also tend to be better and therefore they are better at judging whether a sentence sounds strange and are often capable of fixing an ungrammatical sentence. I therefore asked older children not only to imitate sentences but also to fix each sentence.

Based on the results from the corpus study, I expected children above age 4;6 in both language groups to use the conventional preposition for marking COMPARISON and I did not expect a difference in the rate of use of *from/von* for this construction across languages. For children below age 4;6, the two hypotheses make different predictions. According to the SOURCE HYPOTHESIS, the mean rate of using *von/from* to mark COMPARISON should be the same across languages. However, according to the CONSTRUCTION HYPOTHESIS, young English-speaking children should use *from* more often to mark COMPARISON than German-speaking children use *von*.

4.1 Subjects

The subjects were 17 children acquiring Austrian German and 20 children acquiring American English. The German-speaking children were recruited and tested at the KIWI Alterlaa nursery school in Vienna, Austria, and the English-speaking children were recruited and tested at the Bing nursery school in Stanford, CA. All children were growing up in monolingual households but some of the English-speaking children had Spanish-speaking nannies. I excluded two German-speaking children from my analyses (one boy aged 2;7.4 and one girl aged 3;9.29) as they omitted arguments or responded with completely different sentences more than 80% of the time. I grouped the remaining children into two age groups: younger than 4;6 and older than 4;6. Table 5 shows sex and age statistics for the two age groups in each language. The German and English subject groups slightly differ in two respects. First, the age range of the younger English-speaking group is larger than the one of the younger German-speaking

group as almost all children at the nursery school in Vienna were older than 3;6, and second, the sex distribution varies across languages. While both differences could potentially be problematic, the results from the corpus study do not suggest that there are any differences by sex. Further, all children who used *from* to mark COMPARISON continued to do so up until at least age 3;8, which is older than several of the German-speaking subjects in my sample.

4.2 Procedure

I tested each English-speaking child in a small game room with a table and chairs outside the classroom. Before each session, I put a set of 22 cards, a pig puppet, and a sheet of paper with five colored circles on the table. I took each child into the room and introduced him or her to the pig puppet with the sentences “So, this is Oscar. And do you know what Oscar really likes? When you always say whatever he says. So, for example, if Oscar says *Hello*. What do you say then?”, using a different voice for *Hello*. Most children then replied with “Hello” and otherwise I encouraged them to say “Hello” by saying “Can you also say *Hello*?”. I then said the more complex sentence “*I am a pig with pink ears.*” in Oscar’s voice and again, if children did not immediately imitate, I encouraged them again by saying “Can you also say *I am a pig with pink ears*?”. Once the child imitated the more complex utterance, I said “*The cards are under the table.*” in Oscar’s voice while the cards were on the table, and I encouraged the child again to repeat what Oscar said. Most children imitated the semantically incorrect utterance verbatim but some noticed right away that something was wrong with the utterance. If the child imitated verbatim, I followed up with “But wait, the cards are actually not under the table, no. Oscar sure says silly things sometimes. Can you help him say it better?” and repeated variants of the question “Can you help him say it better?” until the child responded with “The cards are on (top of) the table”. If the child noticed a problem with the utterance right away, I responded with “You are right, the cards are not under the table. Oscar sure says silly things sometimes. Can you help him say it better?”.

After this training session, I explained to the child that we were going to play a short game and explained that the child should draw a card and then Oscar would say a sentence about the card and that the child should say whatever Oscar says. I further explained that after five cards, the child would get a golden star sticker, which he or she could put on the sheet of paper in front of him or her.

The actual session proceeded then as follows. I asked the child to turn over the first card in the pile. Each card contained either a drawing of a child

or a drawing or photo of an object (see Appendix A for the pictures), which corresponded to the subject of a stimulus. I then introduced the person or object on the card with “Oh, that’s ‘NAME’” or “Oh, the ‘OBJECT’” in my voice and then said the sentence associated with the person or the object in Oscar’s voice. When the child did not immediately imitate, I said “Can you also say ----- ?” until the child produced an IMITATION. For the youngest children (aged 3;8 and below¹³), I only asked them to imitate what Oscar said; the other children were also asked whether they could “say it better” to elicit the REPAIR. This procedure was repeated for each of the 22 cards in randomized order. Some of the children were very reluctant to produce repairs and if they repeatedly refused to “say it better”, I reminded them that Oscar sometimes says silly things and that he needs their help.

I recorded all sessions using the Voice Memos app on an iPhone 6.

The procedure for the German children was identical to the described procedure with the exception that a) I used a puppet of the snowman “Olaf” from the Disney movie “Frozen”¹⁴ and b) I tested the children in a corner of the classroom instead of in a separate room. The German script that I used at the beginning of each session is in Appendix C.

4.3 Materials

I tested each child with 22 sentences. Four sentences contained a locative source, four a demoted agent, four a standard of comparison, four an accompanier, and four an instrument¹⁵. Half of the sentences were grammatical and half of them were ungrammatical because they contained a wrong preposition. The remaining two sentences contained a locative source, conventionally marked with *from* in English, and with *aus* (*out of*) in German; these served as fillers to balance the number of prepositions and the number of sentences with *from* that children heard in each language. The English and German sentences were translations of each other with two exceptions. First, I localized the names with the exception of the names of three “Frozen” characters. For both languages, I

13 I originally intended to set the cutoff at 3;5 as Clark & Carpenter (1989b) did. However, some of the children between 3;5 and 3;8 did not produce any repairs and therefore I ended up using a higher cutoff.

14 I intended to use the same puppet at Bing but the research staff recommended that I use something more neutral because “Frozen” is more popular among children in the US than in other countries and they were worried that using a character from the movie might be too distracting for the children.

15 For the question whether German-speaking children mark standards of comparison with *von*, only the four sentences with a standard of comparison are relevant, and all the other items can be seen as fillers. However, I discuss some of the other results in Section 5.

Grammatical sentences	
(8) Olivia went home from school (11) Jacob brought in sand from outside	locative source
(1) Anna got bitten by a large crocodile (4) The monkeys were fed by the children	agent
(5) The mouse is smaller than the bear (6) The elephant is heavier than the rabbit	standard of comparison
(10) Mia played with Emma (13) Hans built a tower with his dad	accompaniment
(14) Sophia caught a duck with her hands (19) James ate the soup with a big spoon	instrument
(15) Noah drank the milk from the bottle (17) The flowers are from the garden	filler
Ungrammatical sentences	
(18) The strawberries are with the farm (22) The bus went with Palo Alto to Stanford	locative source (<i>with</i> → <i>from</i>)
(3) The dog was petted with grandma (12) Lisa was tickled with her mom	agent (<i>with</i> → <i>by</i>)
(7) The plane is faster from the train (9) The horse is bigger from the frog	standard of comparison (<i>from</i> → <i>than</i>)
(2) The cat cuddled from her mom (20) Elsa danced from Dan	accompaniment (<i>from</i> → <i>with</i>)
(16) Emma cut an apple from a knife (21) Ben can fix his toy from a hammer	instrument (<i>from</i> → <i>with</i>)

Table 6 Sentences used in the English elicitation task. The number in front of each sentence denotes the number on the corresponding card in Appendix A. The expected substitutions for the ungrammatical sentences are shown below each category.

only used names or common nicknames of names that were among the 20 most common names for newborns in 2014 in Austria¹⁶ and the US¹⁷, respectively. Second, instead of a translation of the English sentence **The bus went with Palo Alto to Stanford*, I used a German translation of the sentence **The train went with Vienna to Graz* to localize the town names and to give a realistic means of transportation. The list of English sentences is in Table 10 and the German sentences are in Appendix B.

The cards corresponding to the sentences always showed the subject of the sentence. While in theory, the cards were not necessary for the task, their use provided several advantages. First, they made the sentences sound less out of context. By first introducing the person or object on the card, it seemed more natural that the puppet would say something about the person or object as compared to the puppet uttering the sentences out of the blue. Second, for the sentences with names in subject position, the cards provided me with a way to make sure that the children understood that the subject was a person, even if they had not heard that name before. Third, the cards provided a context in which it was natural to promote a patient to subject position as the person on the card became the topic of the conversation. Finally, since the children liked the cards, they made the task more interesting for them.

4.4 Coding

I transcribed all the recordings and then manually coded each of the responses. My main goal was to determine from the responses whether a child would use the conventional (adult-like) preposition for a certain argument.

For the IMITATION responses, I used three coding categories: CORRECT, INCORRECT, and OTHER. CORRECT and INCORRECT responses expressed the correct verb and all its arguments, whereas OTHER responses either omitted some arguments or responded with a completely different sentence. CORRECT responses contained the target preposition whereas INCORRECT responses contained a different preposition. In general, I considered responses with a verb form different from the one in the stimulus (i.e., different in tense or person) as CORRECT or INCORRECT unless the child had changed the response from passive to active or vice versa. For example, I counted substituting *was building* for *built* as CORRECT or INCORRECT, depending on which preposition was used, but I counted substituting *was tickling* for *was tickled* as OTHER. Substituting names, substituting pronouns for noun phrases, swapping the order of noun

16 http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/geborene/haeufigste_vornamen/index.html

17 <https://www.ssa.gov/oact/babynames/>

phrases in comparative constructions, substituting or omitting determiners, and adding or omitting modifiers did not affect the coding decision.

Considering that children provided free form responses and that they interpreted the task in different ways, I elicited some responses that demonstrated their ability to use the right preposition but that would have been coded as OTHER according to my guidelines. These included omitting the sentence-final German auxiliary *werden* ('become') as well as responding with utterances that only contained an oblique noun phrase, as in "*No, with a fork*". I coded these responses as CORRECT or INCORRECT. For the ungrammatical sentences with locative sources, many children changed the preposition to a locative preposition other than *from* (e.g., *the strawberries are at the farm*). As these changes made the sentence grammatical but did not demonstrate whether a child knew how to express locative sources, I coded such responses as OTHER.

For the REPAIR responses, I used the same categories as well as a fourth category, NO RESPONSE, for when a child did not produce a REPAIR response.

Some children frequently made changes to the sentence when I asked them to imitate, but then either repeated what they said before or did not respond to my request for a REPAIR response; other children first imitated verbatim and then changed the preposition in their REPAIR response. To make it easier to compare results, I therefore also combined the IMITATION and REPAIR responses. Generally, I only considered the better response¹⁸ of the two responses unless the child imitated a grammatical sentence verbatim and then substituted a different preposition such that the response became ungrammatical. I considered these cases INCORRECT. For the youngest children who only provided imitations, the COMBINED responses are only determined by (and therefore identical to) their IMITATION responses.

4.5 Results and Discussion

While German and English are very similar in many respects, there are also notable differences between the two languages. For example, German has a more complex case system, is more flexible in terms of word order, and has a more complex gender system than English. All these differences could potentially make the task harder for children in one of the two language groups and therefore invalidate any comparison across languages. Before discussing my results, I therefore first consider whether there were any unexpected differences across the two language groups.

¹⁸ "Better" in this context means to be ranked higher according to the ordering CORRECT > INCORRECT > OTHER > NO RESPONSE.

There was no significant difference in how often German- and English-speaking children were able to produce exact imitations of the stimulus sentences in either age group (younger age group: German 49%, English 61%, $BF=1.2$; older age group: German 66%, English 69%, $BF=0.14$). Second, and more important, if we consider the COMBINED responses and look at how often children made a relevant change to a sentence according to the coding criteria, there was again no significant difference across the two languages (younger age group: German 30%, English 25%, $BF=0.18$; older age group: German 24%, English 27%, $BF=0.25$). The latter result suggests that children in both language groups were equally likely to change a preposition or change an entire sentence, which indicates that children in both language groups were able to perform the task equally well. However, there was one difference between the older German-speaking children and the older English-speaking children: The German-speaking children refused to provide REPAIR responses significantly more often than the English-speaking children (German 74%, English 34%, $BF>10E13$). This implies that the COMBINED responses for the older German-speaking group are significantly more often only determined by their IMITATION responses. Considering that older children tend to be better at imitating sentences verbatim, this difference makes comparisons in terms of how well the older children in both language groups were able to correct ungrammatical sentences problematic because a lower performance by the German-speaking group might solely reflect less willingness to cooperate rather than a lower proficiency¹⁹. For the younger children, there was no significant difference between the two language groups. Overall, these results suggest that the findings from the German and the English data are comparable with the exception of the REPAIR and COMBINED responses to ungrammatical sentences in the older age group.

The overall results are presented in Figure 2 and Table 7. The two plots show the relative frequencies of types of responses to grammatical and ungrammatical sentences for each language and age group. In all age and language groups, the relative frequency of CORRECT responses to ungrammatical stimuli was significantly above 0 ($BF>10E6$ for all groups), which again shows that children are not merely imitating the stimuli or making random changes but instead providing sensible corrections, either in their IMITATION or in their REPAIR responses.

19 I believe there were two reasons for the lower rate of REPAIR responses by the German-speaking children. First, as I was testing the German-speaking children in the classroom, they were more distracted than the English-speaking children, and second, the English-speaking children at BING were much more used to participating in experiments and most likely felt more comfortable doing the task than the German-speaking children did.

Grammatical sentences						
	Younger children			Older children		
	CORRECT	INCORR.	OTHER	CORRECT	INCORR.	OTHER
German	88	0	12	95	2	3
English	84	5	11	92	4	5

Ungrammatical sentences						
	Younger children			Older children		
	CORRECT	INCORR.	OTHER	CORRECT	INCORR.	OTHER
German	41	49	10	43	54	4
English	27	63	10	41	52	7

Table 7 Relative frequencies (%) of types of responses to all grammatical and ungrammatical sentences. The responses are the result of combining the IMITATION and REPAIR responses as described above.

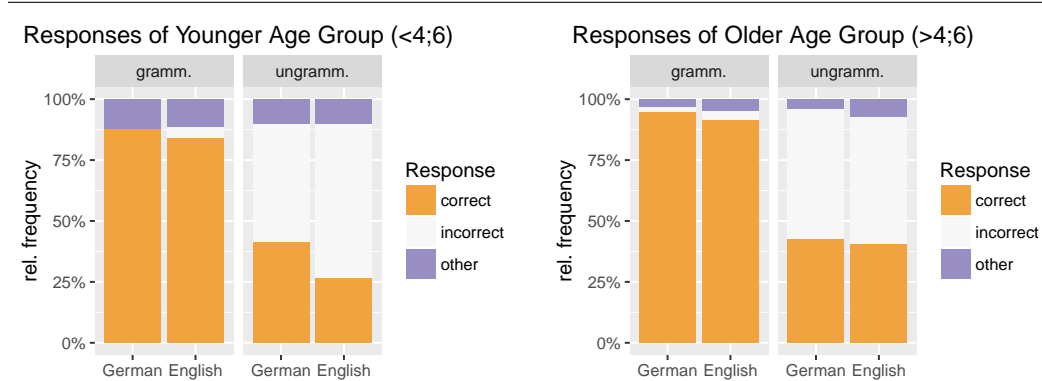


Figure 2 Relative frequencies of types of responses to all grammatical and ungrammatical sentences by younger children (left) and older children (right). The responses are the result of combining the IMITATION and REPAIR responses as described above.

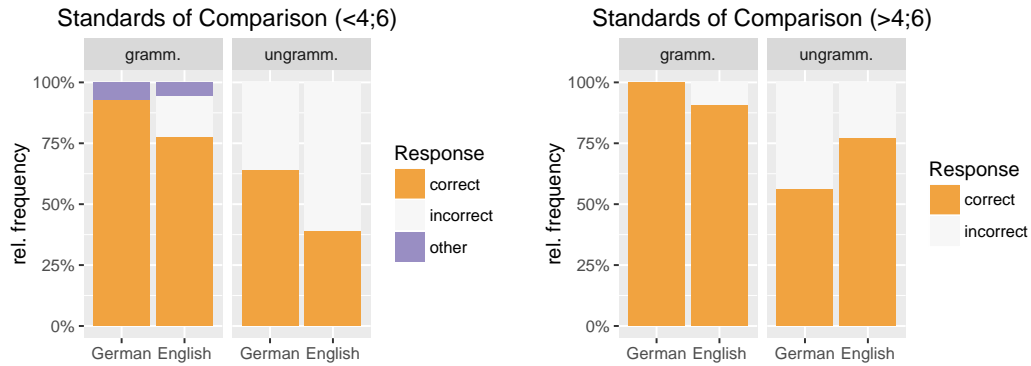


Figure 3 Relative frequencies of types of responses to grammatical and ungrammatical sentences with standards of comparison by younger children (left) and older children (right). The responses are the result of combining the IMITATION and REPAIR responses as described above.

Figure 2 also suggests that English-speaking children in the younger age group were worse at correcting ungrammatical sentences than German-speaking children in the same age group. The primary reason for this difference is that English-speaking children were worse for sentences with COMPARISON and demoted agents. While this difference was not significant ($BF=1.23$), it was very consistent for all stimuli with these two constructions, which suggests that this difference could be significant given a larger sample size.

Figure 3 shows the relative frequencies of responses to the stimuli with COMPARISON. Qualitatively, German-speaking children in the younger age group were better at correcting sentences with *von* marking COMPARISON than English-speaking children in the same age group. Further, only English-speaking children changed grammatical sentences to ungrammatical ones; all of them did so by substituting *from* for *than*. In the older age group, again only English-speaking children changed grammatical sentences. At the same time, older English-speaking children also appeared to be better at substituting *than* for *from* in response to ungrammatical stimuli. Quantitatively, none of these differences were significant. However this is likely a result of the small sample size; notice that these results are very consistent with the corpus results. The only surprising difference is the lower rate of correct responses to the ungrammatical stimuli by the older German-speaking children. This is

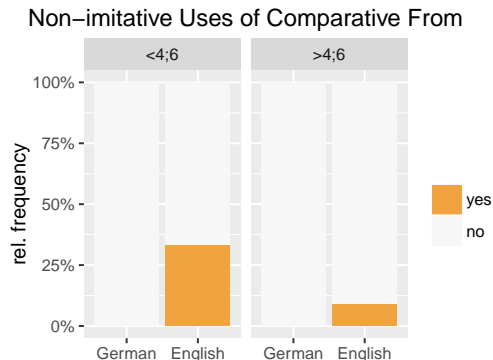


Figure 4 Relative frequencies of children within the younger (left) and older (right) age group who at least once substituted *from/von* for *than/als*.

probably the outcome of the low repair rate in this group, which I mentioned above.

Instead of considering individual responses, we can also consider the entire transcript in a manner similar to that applied to the corpus study. Figure 4 shows the fraction of children within each age group and language who at least once substituted *from/von* for *than/als*. Thirty-three percent of the younger English-speaking children made such a substitution whereas none of the German-speaking children did so. If we incorporate the prior from the corpus study, younger English-speaking children mark COMPARISON significantly more often with FROM (BF=6.32). Among the older children, only one English-speaking child substituted *from* for *than* and the difference between German- and English-speaking children within the older age group was not significant (BF=0.2).

These results are consistent with the results of the corpus study and provide further evidence for the CONSTRUCTION HYPOTHESIS. There is a clear difference between early marking of COMPARISON in German and English, which can be easily explained by the differences in the input. The SOURCE HYPOTHESIS, on the other hand, does not offer an explanation for this difference; both languages express locative source with a preposition and if some English-speaking children use *from* to mark COMPARISON because they perceive the standard of comparison as a source, then some German-speaking children should show a similar behavior.

From a qualitative point of view, the results of the elicitation task further highlight the importance of the adult input. One German-speaking child

	big	good	old	different	bad	tall	small
Overall frequency	1,407	2,493	338	190	263	78	173
Comparative constructions	32	27	12	9 (<i>from</i>) 3 (<i>than</i>)	9	8	5

Table 8 Overall frequency and frequency of comparative constructions with an explicit standard of comparison for several adjectives in the English adult input, summed across the corpora from Section 3. These are all the adjectives that appear at least 5 times in a comparative construction with a standard of comparison in the adult input.

(F;3.11.0) substituted *wie*, the equative-marking preposition, for *von*. As I mentioned earlier, *wie* is used by some adults to mark COMPARISON.

I conclude from the results of the corpus study and the elicitation task that English-speaking children most likely overextend the construction *different from* rather than relying on a conceptual category of SOURCE when they mark COMPARISON with *from*. However, this conclusion raises another question: Why do comparative constructions with *different* play such a prominent role? Several other adjectives such as *good*, *big*, *nice*, and *old* appear much more frequently in the adult input and therefore it seems surprising that the construction *different from* has a significant impact on children’s early marking of standards of comparison. However, this becomes less surprising if we consider only comparative adjectives with an explicitly expressed standard of comparison because the construction *different from* is among the most frequent comparative constructions in the adult input. Table 8 shows the frequencies of adjectives that appear at least 5 times in comparative constructions in the combined adult input of the corpora from the previous section. As this table shows, despite the fact that the adjective *different* appears overall much less frequently than *big* or *good*, the construction *different from* is among the most frequent comparative constructions. This table also shows that adults rarely express standards of comparisons in child-directed speech; note that these numbers are the total uses in 6 longitudinal corpora. This observation suggests that on the one hand, a small number of adult utterances with *different from* could have a significant impact on a child’s acquisition of comparative constructions, and on the other hand, it suggests that if a child assumes that *from* can be used to mark COMPARISON, it might take him or her a long time to acquire the conventional means as he or she receives little input that contradicts the child’s hypothesis.

5 Demoted Agents and Possessors

In the previous two sections, I argued that the non-conventional uses of *from* to mark COMPARISON are most likely not a result of children perceiving the prepositional complement as a source. However, my account does not provide an explanation for the non-conventional marking of demoted agents or possessors, which Clark & Carpenter (1989a,b) observed in corpora and – in the case of demoted agents – in an elicitation task. Considering that German-speaking children as well as adults also use *von* to mark possessors and demoted agents, all the data are compatible with the following REVISED SOURCE account.

Children perceive temporal sources, demoted agents, causes, natural forces, and possessors (but not standards of comparison) as a source and therefore mark all of these arguments with the preposition *from* or the equivalent marker of locative sources in other languages.

However, there are some observations that while not contradicting the REVISED SOURCE account, raise the question whether it is the best explanation for the data. First, if we assume a relationship between conceptual categories and language typology, then the REVISED SOURCE category seems very unlikely. There are only three languages in a sample of 98 languages that show the case syncretism of REVISED SOURCE: German (using *von*), Middle English (*of(f)*), and Old Spanish (*de*) (Palancar 2002). If REVISED SOURCE was a prominent and universal conceptual category, we might assume that more languages could express all these arguments with the same means.

Second, Clark & Carpenter (1989a) argue that all English-speaking children first assign the meaning of a locative source to *from* based on the fact that the first utterance with *from* of all investigated children expressed a locative source, and they argue that the non-conventional uses are an overextension of the locative meaning. However, this is not the case for German-speaking children. Only two of them marked a locative source in their first utterance with *von*, whereas the other four either marked a possessor or used a partitive construction.

Third, the REVISED SOURCE account no longer assumes that children assign only one meaning to *from* but rather that some children assume that *from* marks SOURCE as well as COMPARISON. While children might still initially assume that *from* is monosemous and can only be used to mark SOURCE, some children start to use *from* to mark COMPARISON before age 3 and from that point on they must assume that *from* is polysemous. This is per se not problematic but once we abandon the assumption that children assign only

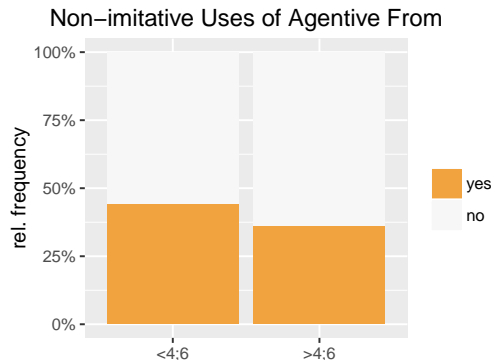


Figure 5 Relative frequencies of English-speaking children within the younger (left) and older (right) age group who at least once marked an agent with *from*.

one meaning to *from*, there is no reason to believe that children cannot assign three or more meanings to *from*.

Fourth, possessors and demoted agents vastly differ in frequency – both in the adult input and in early productions. Demoted agents are rare in spoken language whereas possessors are very frequent. Nevertheless, English-speaking children use *from* approximately as often to mark demoted agents as they use it to mark possessors. At the same time, they use *'s* and *of* two orders of magnitude more often than they use *from* to mark possessors.

All of these observations suggest that there might be a better explanation for the non-conventional uses of *from* for demoted agents and possessors. I discuss these types of arguments in detail in turn.

5.1 Demoted agents

The use of *from* for demoted agents in English appears to be the most consistent non-conventional use of *from*. In my elicitation task, I was able to reproduce the results by Clark & Carpenter (1989b). Figure 5 shows the relative frequency of children who marked agents with *from* for both age groups. Forty-four percent of the younger children and 36% of the older children used *from* at least once to mark an agent despite the fact that all of the stimuli with demoted agents marked them using either *by* or *with*. The relative frequencies in both age groups are significantly greater than 0 (younger age group: BF=4.66, older age group: BF=3.54).

These non-conventional uses are particularly interesting because adults only infrequently use demoted agents in conversational speech and for this reason, demoted agents are virtually non-existent in the input. On average, adult utterances with a demoted agent appeared only 9 times within each corpus and this number already includes several out-of-context uses in the Adam, Eve, and Sarah corpora to test whether the child was able to understand passive constructions. Considering that children rarely hear demoted agents marked with *by*, it is therefore not surprising that children do not know the conventional means to express demoted agents. Nevertheless, the lack of input does not explain why children consistently opt for *from*.

One explanation for this behavior is provided by the REVISED SOURCE account. Clark & Carpenter (1989a) argued that children perceive agents as the source of an event and therefore mark demoted agents using the preposition that is used to mark locative sources. However, there are two other possible explanations, which do not assume that children make abstract overgeneralizations.

First, it could be that some children perceive demoted agents as a cause and therefore mark them with *from*. Some evidence for this hypothesis comes from the typological literature. In the sample of 98 languages by Palancar (2002), 46 languages from a large variety of families mark demoted agents and causes using the same means. This suggests that this case syncretism developed independently, which in return might suggest that there could be a cognitive basis for perceiving demoted agents as causes. Further evidence comes from the findings that children tend to use adjectival *get*-passives before regular passives (Borer & Wexler 1987, Fox & Grodzinsky 1998), which could suggest the following developmental path. Children first acquire the conventional construction *get ADJ from CAUSE* and then overextend this construction to the non-conventional *get PARTICIPLE from AGENT* and *be PARTICIPLE from AGENT* before realizing that the conventional construction is *be/get PARTICIPLE by AGENT*. None of the English corpora that I investigated are dense enough to confirm or disprove this developmental path, but some children did produce utterances with the construction *get ADJ from CAUSE* at an early age (e.g., *Who gets sick from eating seeds?*, Abe, 3;1.8), and all other constructions appeared in some responses to the elicitation task.

Second, it could be that children overextend the use of *from* to mark natural forces. While this also appears to be a plausible overextension, this hypothesis is very unlikely to be true given the fact this construction appears even less frequent in the adult input than demoted agents. Further, apart from Damon, only Sarah used *from* to mark a natural force and she did so after she had begun to mark agents with *from*.

	Adam	Sarah	Eve	Shem	Abe	Naomi
<i>'s</i>	272	179	60	203	446	184
<i>of</i>	592	268	70	160	782	79
<i>from</i>	0	1	0	5	7	0

Table 9 Frequencies of possessive *'s*, *of*, and *from* used to mark possessors.

It could also be that children take all these constructions as evidence that agents should be marked with *from*. The two hypotheses do not contradict each other and therefore it could be that all of these constructions contribute to children’s assumptions on how agents should be marked.

All of this is speculative at this point, and I do not have additional evidence for either hypothesis. However, there seems to be as much evidence that children perceive demoted agents as causes, as there is that they perceive demoted agents as sources. Further, both of my explanations assume that less abstraction is required of the child and therefore I argue that they should be considered in future work on this topic.

5.2 Possessors and Partitives

The phenomenon that children mark possessors with *from* appears to be quite distinct from the non-conventional expression of demoted agents for several reasons. First, exchange and possession are among the earliest and most frequent topics of conversation of young children across many different cultures (e.g., Stern & Stern (1928), Brown (1973), Tomasello (1998)). Second, children know the conventional means of expressing possession: All children who marked possessors with *from* expressed possession with the clitic *'s* or the preposition *of* at least two orders of magnitude more often than they did with *from* (see Table 9) and all children talked about possession before or at the same time as they produced their first utterance with *from* to mark a locative source. Children at any age therefore clearly do not assume that *from* is the primary device to mark possessors but at most that it can be used along with *'s* and *of*.

Some of the non-conventional uses appear to be overextensions of conventional constructions. For example, Shem produced the utterance *dat's [that's] a finger from him* (3;0.13) a few days after he uttered *dat's [that's] from my – um – puppet show* (3;0.5, talking about a lion puppet). Both utterances express a part-of relation between two objects: The puppet is a part of the puppet show and the finger is part of the person that Shem is talking about. However, the construction *that's NP1 from NP2* can only be used when NP1 can be

physically detached from NP2, which makes *that's a finger from him* ungrammatical. This difference might not be completely apparent to a three-year-old child and it is likely that he overextended the conventional construction.

Other non-conventional uses such as *I see boats from mommy* (Adam, 3;0.0) or *you can be a mom from two babies* (Damon, 3;7.5) are harder to explain based on conventional constructions and they might be rooted in non-linguistic conceptual difficulties. Despite the fact that children talk a lot about possession, they have difficulty in determining ownership of objects up until at least age 5. Friedman & Neary (2008) ran a study that showed that children up until age 5 have a strong bias towards assigning ownership of a toy to whoever interacted with it first. This bias persists to a large extent even if the person who first interacted with the toy gives the toy to another person and the child is being told that the second person got the toy as a present. This bias can only be overcome if the presentation is made more prototypical by wrapping the toy and not letting the first person interact with it. It is therefore conceivable that children misinterpret frequent adult utterances such as *That's a present from Granny Hart* (Mother to Eve, 1;9) to mean that the grandmother still owns the toy rather than that the child got the toy from the grandmother. If children misinterpret such utterances, then it would not be surprising if they occasionally used *from* to mark possessors.

These explanations are again speculative but they again illustrate that there exist reasonable explanations that do not assume that children perceive possessors as an abstract source.

6 Conclusion and Future Directions

In this paper, I revisited the phenomenon that English-speaking children use the preposition *from* for more types of arguments than adults. I compared how English- and German-speaking children express standards of comparison in a corpus study as well as an elicitation experiment, and I argued on the basis of the results from these two studies that the non-conventional uses of *from* to mark standards of comparison are the result of overextending the conventional construction *different from*, one comparative form that children hear from early on.

I also discussed the uses of *from* to mark demoted agents as well as possessors and presented new explanations for both phenomena that do not assume that children can make very abstract generalizations. These explanations, however, are only speculative at this point and more work is needed to determine whether any of these possibilities are superior to the REVISED SOURCE account. One possible avenue to test some of these hypotheses would be to investigate the use

of prepositions in other languages with different case syncretisms. In Italian, for example, adults mark demoted agents with the source-marking preposition *da* but they mark causes and possessors with *di*. Therefore, the presence or absence of consistent non-conventional uses of either preposition could provide new insights on why English-speaking children mark demoted agents with *from*.

To conclude, I argued for an explanation of some of the non-conventional uses of *from* that does not assume that children rely on an abstract conceptual category in formulating their utterances. While this work clearly falls short of being a comprehensive account of the acquisition of grammatical prepositions, it demonstrates the complexity of this process and shows that future research in this area should focus more on specific constructions in different languages rather than only focusing on what all non-conventional uses of a grammatical preposition have in common.

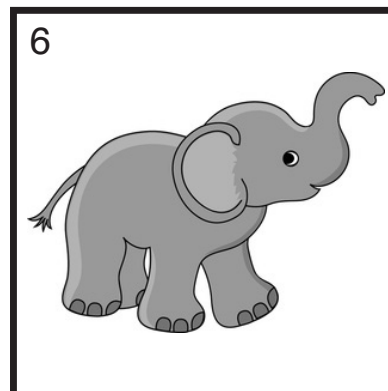
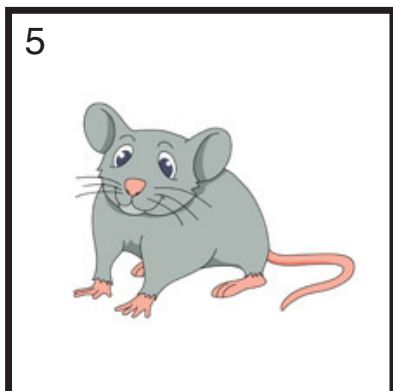
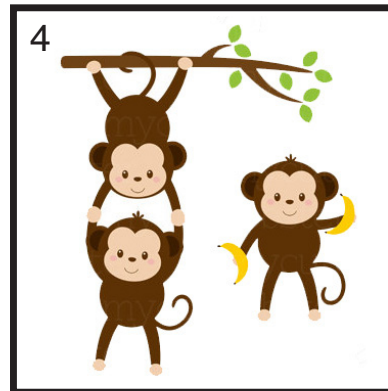
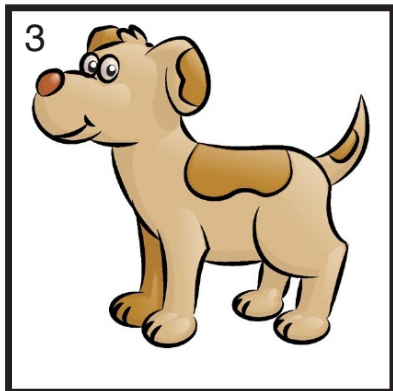
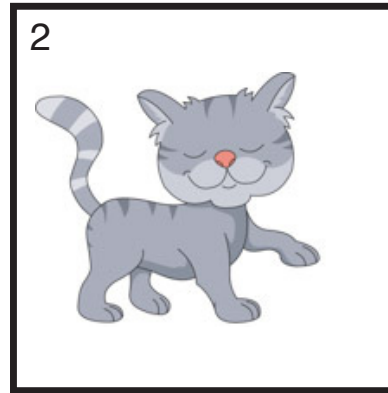
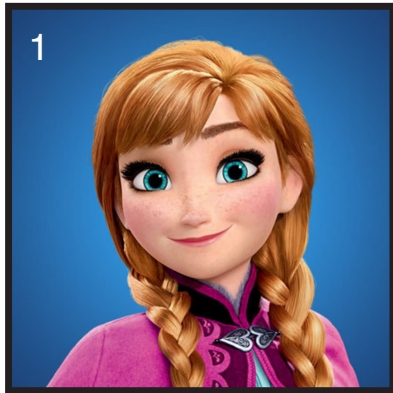
References

- Barker, Chris. 1995. *Possessive Descriptions*. Stanford, CA: CSLI Publications.
- Behrens, Heike. 2006. The input–output relationship in first language acquisition. *Language and Cognitive Processes* 21(1-3). 2–24.
- Borer, Hagit & Kenneth Wexler. 1987. The maturation of syntax. In Thomas Roeper & Edwin Williams (eds.), *Parameter Setting (Studies in Theoretical Psycholinguistic)*, 123–172. Dordrecht: Springer Netherlands.
- Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Chouinard, Michelle M. & Eve V. Clark. 2003. Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30(3). 637–669.
- Clark, Eve V. 1978. Awareness of language: Some evidence from what children say and do. In Anne Sinclair, Robert J. Jarvella & Willem J. M. Levelt (eds.), *The Child's Conception of Language*, 17–43. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Clark, Eve V. & Kathie L. Carpenter. 1989a. The notion of source in language acquisition. *Language* 65(1). 1–30.
- Clark, Eve V. & Kathie L. Carpenter. 1989b. On children's uses of *from*, *by* and *with* in oblique noun phrases. *Journal of Child Language* 16(2). 349–364.
- Efron, Bradley. 2005. Bayesians, Frequentists, and Scientists. *Journal of the American Statistical Association* 100(469). 1–5.
- Fox, Danny & Yosef Grodzinsky. 1998. Children's Passive: A View from the By -Phrase. *Linguistic Inquiry* 29(2). 311–332.
- Friedman, Ori & Karen R. Neary. 2008. Determining who owns what: Do children infer ownership from first possession? *Cognition* 107(3). 829–849.

- Goodman, Noah D & Andreas Stuhlmüller. 2014. The design and implementation of probabilistic programming languages. <http://dippl.org>. Accessed: 2016-8-31.
- Hiller, Sarah & Raquel Fernández. 2016. A data-driven investigation of corrective feedback on subject omission errors in first language acquisition. In *Proceedings of the SIGNLL conference on computational natural language learning (CoNLL)*, 105–114. Association for Computational Linguistics.
- Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge, MA: MIT Press.
- Jeffreys, Harold. 1961. *Theory of Probability*. Oxford: Oxford University Press 3rd edn.
- Johnson, Christopher. 1996. Learnability in the acquisition of multiple senses: SOURCE reconsidered. *Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Role of Learnability in Grammatical Theory* 469–480.
- Kidd, Evan & Thea Cameron-Faulkner. 2008. The acquisition of the multiple senses of with. *Linguistics* 46(1). 33–61.
- Kuczaj, Stan A. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16(5). 589–600.
- Lee, Michael D. & Eric-Jan Wagenmakers. 2013. *Bayesian Cognitive Modeling*. Cambridge: Cambridge University Press.
- MacWhinney, Brian. 2000. *The CHILDES project: The database*, vol. 2. Psychology Press.
- Maienborn, Claudia & Johanna Herdtfelder. 2015. A compositional account of the eventive/stative ambiguity of German causal von-modifiers. In Sarah D’Antonio, Mary Moroney & Carol R. Little (eds.), *Proceedings of the 25th semantics and linguistic theory conference*, 163–183. Linguistic Society of America and Cornell Linguistics Circle.
- McKercher, David A. 2001. *The polysemy of WITH in first language acquisition*: Stanford University dissertation.
- Miller, Max. 1979. *The Logic of Language Development in Early Childhood*, vol. 3 Springer Series in Language and Communication. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Palancar, Enrique L. 2002. *The Origin of Agent Markers*. Berlin: Akademie Verlag.
- Sachs, Jacqueline. 1983. Talking about the there and then: The emergence of displaced reference in parent-child discourse. In Keith E. Nelson (ed.), *Children’s language*, vol. 4, 1–28. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Slobin, Dan I. & Charles A. Welsh. 1973. Elicited Imitation as a Research Tool in Developmental Psycholinguistics. In Charles A. Ferguson & Dan I. Slobin (eds.), *Studies of child language development*, New York, NY: Holt,

- Rinehart and Winston.
- Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford, England: Blackwell.
- Stern, Clara & William Stern. 1928. *Die Kindersprache*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- von Stuckrad, Gesine. 1957. Denn – dann in historischer Sicht vom Althochdeutschen bis zum Neuhochdeutschen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* (79). 489–535.
- Thurmair, Maria. 2001. *Vergleiche und Vergleichen*. Berlin, Boston: De Gruyter.
- Tomasello, Michael. 1987. Learning to use prepositions: a case study. *Journal of Child Language* 14(01). 79.
- Tomasello, Michael. 1998. On child's early talk about possession. In John Newman (ed.), *Linguistics of Giving*, vol. 36, 349–373. John Benjamins Publishing Company.
- Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74(3). 209–253.
- Wolff, Phillip. 2003. Direct causation in the linguistic coding and individuation of causal events. *Cognition* 88(1). 1–48.

Appendix A: Pictures for Elicitation Study



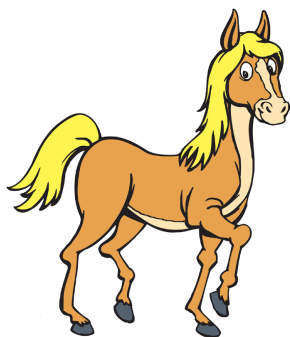
7



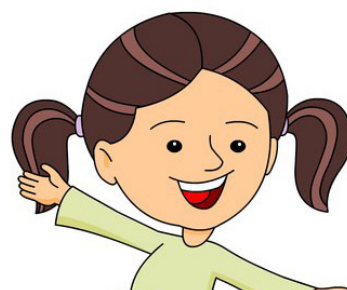
8



9



10

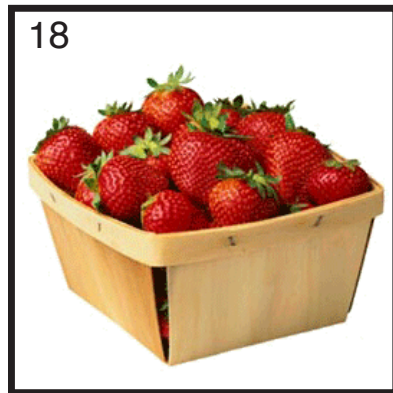
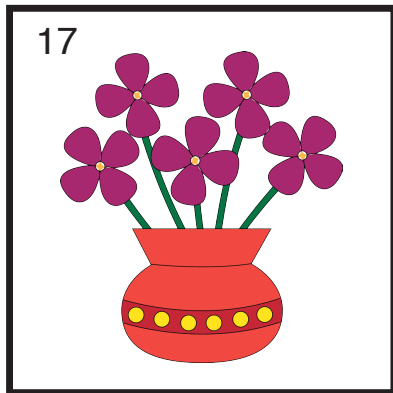
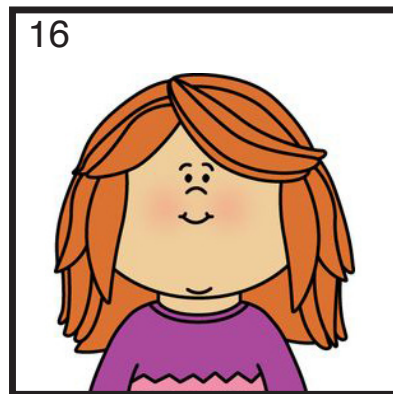
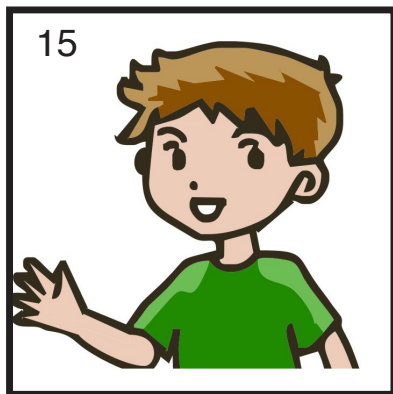


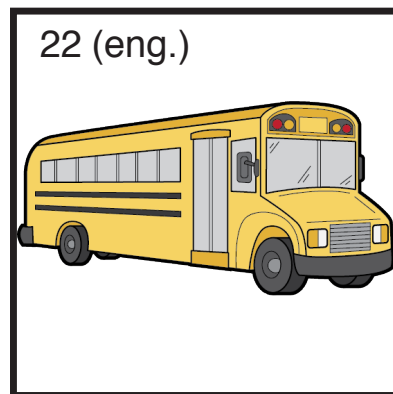
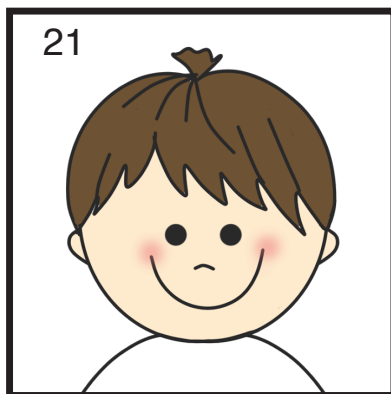
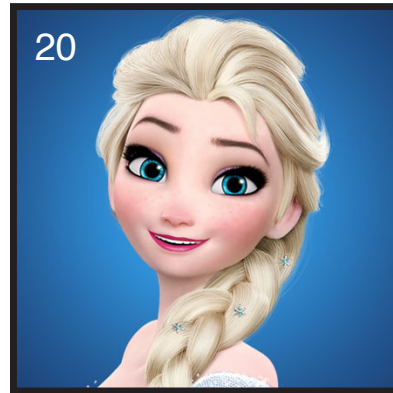
11



12







Appendix B: German Stimuli

Grammatical sentences	
(8) Sarah ist von der Schule nachhause gegangen (11) Tim hat den Sand von drauen mitgebracht	locative source
(1) Anna ist von einem groen Krokodil gebissen worden (4) Die Affen sind von den Kindern gefuttern worden	agent
(5) Die Maus ist kleiner als der Br (6) Der Elefant ist schwerer als der Hase	comparison
(10) Mia hat mit Emma gespielt (13) Hans hat mit seinem Papa einen Turm gebaut	accompaniment
(14) Marie hat eine Ente mit ihren Hnden gefangen (19) Max hat die Suppe mit einem groen Lffel gegessen	instrument
(15) Lukas hat die Milch aus der Flasche getrunken (17) Die Blumen sind aus dem Garten	filler
Ungrammatical sentences	
(18) Die Erdbeeren sind mit dem Bauernhof (22) Der Zug ist mit Wien nach Linz gefahren	locative source (<i>mit</i> → <i>von</i>)
(3) Der Hund ist mit Oma gestreichelt worden (12) Lisa ist mit ihrer Mama gekitzelt worden	agent (<i>mit</i> → <i>von</i>)
(7) Das Flugzeug ist schneller von dem Zug (9) Das Pferd ist grer von dem Frosch	comparison (<i>von</i> → <i>als</i>)
(2) Die Katze hat von ihrer Mama gekuschelt (20) Elsa hat von Elias getanzt	accompaniment (<i>von</i> → <i>mit</i>)
(16) Julia hat einen Apfel von einem Messer geschnitten (21) Felix kann sein Spielzeug von einem Hammer reparieren	instrument (<i>von</i> → <i>mit</i>)

Table 10 Sentences used in the German elicitation task. The number in front of each sentence denotes the number on the corresponding card in Appendix A.

Appendix C: German Script

(Passages in *italics* were uttered in “the puppet’s” voice.)

E(xperimenter): Das is Olaf! Weißt du was Olaf ganz gerne mag? Wenn man immer genau das sagt, was er sagt. Also wenn er sagt “*Hallo*”, kannst du dann auch “Hallo” sagen?

C(hild): Hallo

E: *Ich habe eine Karotte als Nase.*

C: Ich habe eine Karotte als Nase.

E: *Die Karten liegen unter dem Tisch.*

C: Die Karten liegen unter dem Tisch.

E: Aber das stimmt ja gar nicht, oder? Der Olaf sagt manchmal dumme Sachen. Kannst du ihm helfen das besser zu sagen?

C: Die Karten liegen auf dem Tisch.