



The Acquisition Sketch Project

Exploring case marking in German first language acquisition using the acquisition sketch approach

Gianna Urbanczik

University of Cologne, University of Potsdam, Macquarie University

Abstract

This paper provides an example of a sketch description as suggested in the Sketch Acquisition Manual (SAM) in this issue. The SAM provides guidelines for sketch descriptions of child language and child-directed language in under-studied languages. Since this sketch approach has not been tested for well-researched languages yet, the aim of this paper was to evaluate this novel method. Therefore, five hours of natural conversation data from two children acquiring German as their first language were processed according to the guidelines proposed in the manual. The analysis focused on the acquisition of German case-marked definite articles as there is a large body of research in this area. Hypotheses about German first language acquisition based on the identified patterns were then compared to existing literature on the acquisition of German grounded in a large amount of data. All findings were consistent with the existing (case) acquisition literature. However, there were also phenomena which are described in the literature but were not attested by the sketch analysis. The results suggest that the sketch format cannot be expected to reveal all language acquisition phenomena of a certain language but may have the potential to uncover the most evident acquisition processes.

Keywords: language acquisition, child language, case marking, definite articles, German, corpus research

1 Introduction

The present paper contributes an evaluation of the sketch approach developed by Defina et al. in the Sketch Acquisition Manual (the SAM, Parts I and II; 2023a, 2023b). This approach aims to guide researchers in the construction of small, comparable corpora of child language, thus contributing to reducing the typological gap in language acquisition research (see Kidd & Garcia 2022). The SAM provides principles for describing child language and child-directed language in under-researched languages in sketch form.

This paper is the first test of this approach for a well-described language. For this purpose, I identified and processed existing data of German language acquisition according to the guidelines proposed in the SAM. Given the exemplary and evaluating nature of the following sketch on German, I will focus on only one aspect of first language acquisition in German: the acquisition of case.

German was chosen to put the sketch approach to the test because it is a well-documented language in general, but also a well-studied language in terms of language development. For the sake of brevity, I selected a single linguistic phenomenon of the German language: the marking of case on definite articles. There were three main reasons for picking this phenomenon: 1) it is likely to occur frequently in spontaneous speech; 2) we can expect a developmental trajectory in children between 2;0 and 4;0 (YEAR;MONTH), which is the age range predefined by the SAM (Part I); 3) case marking has been extensively described and studied over the course of the acquisition (e.g., Clahsen 1984; Mills 1985; Tracy 1986; Clahsen et al. 1994; Eisenbeiß 2003; Bittner 2006; Eisenbeiß et al. 2006; Ulrich et al. 2016; Scherger 2018; Ulrich & Motsch 2018; Szagun 2019).

The following sections provide an overview of the sketch data used. Subsequently, a sketch of the acquisition of case in German following the guidelines given in the SAM is presented. As part of the sketch, I derive hypotheses about case acquisition in German. Following the sketch, I evaluate the approach proposed by Defina and colleagues (2023a, 2023b) by exploring whether the constructed hypotheses are confirmed or disconfirmed when compared to literature on the acquisition of German grounded in a large amount of data.

Note that the approach of this paper differs both from the regular SAM procedure and from common first language acquisition studies. Unlike in typical SAM studies, this paper investigates one specific phenomenon for which previous acquisition studies are available. First language acquisition studies typically begin by reviewing previous relevant literature, then develop hypotheses based on this literature, and finally present new data to test the hypotheses. The present paper follows the opposite structure, first presenting already existing data, then developing hypotheses based on this data, and finally reviewing previous literature.

2 Corpus construction

Data in the form of acquisition corpora was sought to meet the criteria given in the corpus construction chapter in Part I of the SAM as closely as possible, as will be detailed below. To this end, I screened all German corpora available on the Child Language Data Exchange System (CHILDES – grant support: NICHD HD082736; MacWhinney 2000) of unimpaired acquisition, applying three inclusion criteria: the corpus must 1) cover the desired range of 2;0 to 4;0; 2) provide audio files; and 3) consist of sessions with a minimum duration of 30 minutes. This procedure left two corpora, the Leo corpus (Behrens 2006) and the Rigol corpus (Lieven & Stoll 2013), which thus served as the basis for the sketch.

2.1 Structure of the sketch corpus

The sketch corpus consists of data from two children, Leo and Pauline, extracted from the Leo corpus and the Rigol corpus respectively. The Leo corpus consists of around 380 60-minute audio recorded sessions. The focal child is Leo, who was 1;11 at the beginning of the study and was frequently recorded up to the age of 4;11. The Rigol corpus consists of around 1900 30-minute recordings from 21 children, though only data from four children is available on CHILDES. Out of these four, Pauline was selected. The recordings started shortly after birth or at the age of one and ended in the eighth year of life. In sum, five 30-minute audio recordings of each child at the ages 2,0, 2;6, 3;0, 3;6 and 4;0 form the present sketch corpus.

2.1.1 The children

Both Leo and Pauline grew up monolingually with German as their first language. Leo lives in Leipzig, Germany, with his parents and his sister, who was born when he was aged 3;3. Both parents speak standard High German. Leo's mother has a background in bookselling, his father works in academia. Further background information is provided by Behrens (2006). Pauline lives with her parents and her brother, who is three years older than her. Both parents completed higher education (university). Background information on the place of residence is not publicly available for privacy reasons.

2.1.2 The sessions

A total amount of five hours of data was analyzed. Five sessions for each child were selected from the corpora. Approximately 30 minutes of each recording was analyzed, which lead to one hour of data per time point (see Table 1).

Sessions were chosen based on the following criteria. Firstly, sessions were selected to correspond as closely as possible with the desired age at date of recording. Secondly, the goal was to include sessions with various family members or other persons who play a role in the child's social life, while excluding the participation of investigators. This was done to identify and work with data where settings were as natural as possible, and where the child encountered various interlocutors, resembling a recording in a field situation. The participation of an investigator could only be avoided in the Leo corpus. The investigator was present during every session in the Rigol corpus.

30 minutes per recording were chosen for further analysis primarily based on the child's talkativeness throughout the session. In one case, the recording was shorter than the required 30 minutes (recording of Pauline at the age of 2;6). An additional two minutes were therefore taken from the corresponding session of the other corpus (see Table 1). Detailed information on which time frames were analyzed from which session as well as the exact session names as available on CHILDES (MacWhinney 2000) can be found in the Appendix.

2;0 Age 2;6 3;0 3;6 4:0 Leo (Leo corpus) 30(65) 32(60) 30(60) 30(60) 30(63) Pauline (Rigol corpus) 30(31) 28(28) 30(37) 30(32) 30(44)

Table 1. Session length at each time point.

60(96)

Total

Note. The first number is the number of minutes analyzed per session. The number inside the parentheses is the duration of the entire session in minutes. Adapted from Defina et al., this issue, Part I, Table 1.

60(97)

60(92)

60(107)

60(88)

In Leo's case, all sessions took place at his home and were recorded between 1999 and 2001. During the recordings, Leo engages with one or both of his parents. Sometimes other family members are present during the recordings as well. Leo is fascinated by trains and transportation in general, talking about these topics frequently, and enjoys playing with toy trains and cars (see Table 2). The included sessions with Pauline were recorded between 1993 and 1995. Recordings took place either at the investigator's or at Pauline's home. Detailed information on where each session was recorded can be found in the Appendix. Both the investigator and Pauline's brother were present during every session. The mother joined four

of the five sessions. From Pauline's activities during the five recordings, it appears that she does not have a particular favorite activity, but rather enjoys a variety of activities (see Table 2).

Table 2. Details on each session.

		Participants ^b						
	Exact age ^a	Adults	Children (Age ^a)	Content				
	2;00.06	MOT		Playing with building bricks, building houses, tunnels, and bridges, naming colors, playing with cars and trains, crafting				
	2;06.12	FAT, MOT		Playing with trains, building train tracks, playing mechanic				
Leo	3;00.00	FAT, GMOT on the phone		Playing with cars, talking about ships and buses				
	3;06.22	MOT, GMOT		oking at a map, speaking about various aces in the world and travelling a chat, asking lots of questions,				
	4;00.18	MOT	SIS (ca. 0;09.00)	Having a chat, asking lots of questions, doing a jigsaw puzzle				
	2;00.05	INV	BRO (5;01.28)	Playing with cars, watching a cat, looking at books; P. comments on everything she does				
	2;06.13	INV, MOT	BRO (5;08.06)	Speaking about being recorded, doing a jigsaw puzzle, playing with LEGO duplo, looking at picture books, doing gymnastics, fighting with brother				
Pauline	3;00.11	INV, MOT BRO (6;02.01)		Playing a board game, naming the rooms of a doll's house and explaining her decision, playing with cards				
	3;06.12	INV, MOT	BRO (6;08.05)	Playing and fighting with brother, hiding and seeking objects, wrapping presents, drawing a picture				
	4;00.07	INV, MOT	BRO (7;02.00)	Playing with a soft toy, pretend play cooking and serving the cooked food				

Note. MOT = Mother, FAT = Father, GMOT = Grandmother, INV = Investigator, SIS = Sister, BRO = Brother; aY;MM.DD; bOnly the participants present during the analyzed time frames of the session are listed.

2.2 Data processing and approach to analysis

Data analysis was performed solely on the basis of the transcripts provided for the sessions, which were downloaded from CHILDES (MacWhinney 2000). I compared the transcriptions and the audio files to ensure accuracy of the transcripts. After download, each transcript was imported into an ELAN (2020; Brugman & Russel 2004) file along with the corresponding audio file. All utterances of the focal child were exported from the ELAN file as a text

document and imported into a Microsoft Excel file. Thereafter, each utterance produced within the relevant 30 minutes of the session was analyzed regarding the use of definite articles.

3 Sketch

The primary goal of the present paper is to evaluate the sketch grammar model rather than to provide a sketch of an already well-studied language. The aim, therefore, is to assess whether the novel sketch format enables researchers to sketch possible developmental trajectories and whether it has the potential to serve as a basis for well-founded hypotheses for further language acquisition research. This sketch is thus limited to the investigation of one aspect of morphosyntax (i.e., case marking on definite articles) throughout the course of acquisition of a well-studied language (i.e., German). It will provide a short overview of the German case system and will then focus on a qualitative description of case marking in child language. These parts are guided by the sections on morphology and syntax of Part II of the SAM. For the evaluation, the validity of the derived hypotheses is compared with existing acquisition literature (see Section 4).

3.1 German case system

The German language has four cases: the nominative, the genitive, the dative, and the accusative. Case is marked on articles (both definite and indefinite articles), pronouns, (attributive) adjectives, and, in some instances, on the noun (Hentschel & Weydt 2021).¹

Case serves as a marker of syntactic relations, with the nominative used for subjects and the accusative and the dative (and in very few instances the genitive) marking objects (Eisenberg 2020b: 48-49). An additional function of case is the marking of semantic relations (Eisenberg 2020b: 83).

In German, verbs (e.g., der Nachbarin_[DAT] helfen '(to) help the neighbor'), adjectives (des Betrugs_[GEN] verdächtig 'suspected of fraud'), and prepositions (mit der Tasche_[DAT] 'with the bag') can assign case (Hentschel & Weydt 2021). In a prepositional phrase (PP), the preposition governs the case of its complement (typically a noun phrase (NP); Eisenberg 2020b: 199–200). Prepositions can assign three cases: the accusative, the dative, and the genitive (Meibauer et al. 2015: 132; Hentschel & Weydt 2021: 268–271). Some prepositions can assign two cases (i.e., two-way prepositions). A two-way preposition (e.g., auf) assigning the accusative (e.g., Ich laufe auf den Steg_[ACC]. 'I am walking onto the jetty.') yields a directional meaning, while a two-way preposition assigning the dative (e.g., Ich laufe auf dem Steg_[DAT]. 'I am walking on the jetty.') yields a static-local meaning (Gutzmann & Turgay 2011: 172; Hentschel & Weydt 2021: 271).

¹ Abbreviations: ACC – accusative; ART – article; DAT – dative; DEF – definite; F – feminine; GEN – genitive; M – masculine; N – neuter; NOM – nominative; PL – plural; PTC – particle; SG – singular.

		Singular		Plural
	Masculine	Feminine	Neuter	Masculine/Feminine/Neuter
Nominative	der	die	das	die
Genitive	des	der	des	der
Dative	dem	der	dem	den
Accusative	den	die	das	die

Table 3. Paradigm of definite articles in German.

Case marking requires the specification of number and, in most instances, gender. For definite articles, the paradigm distinguishes the four cases, masculine, feminine, and neuter forms, as well as singular and plural forms (see Table 3). Articles must agree with the noun they determine in case, number, and gender. While there is no distinction of gender in plural contexts, marking of the inherent gender of the noun is mandatory in singular contexts. As can be seen in Table 3, distinctive marking of all four cases solely occurs in definite articles of masculine nouns (in the singular). Furthermore, the articles *des* and *dem* are the only articles for which an unambiguous assignment of case is possible. Consequently, due to the high number of syncretisms, errors in the production of definite articles cannot always be clearly attributed to one cause.

(1) *mit *der Ball* 'with the ball'

If, for instance, a child uttered this phrase (1), above, instead of *mit dem Ball* 'with the ball', the error could be attributed to the production of an erroneous case, i.e., the production of nominative instead of dative, or equally an error of gender if the child treated *Ball* 'ball' as a feminine instead of a masculine noun.

Approach to analysis: Inclusion and exclusion criteria

The goals of the analysis were identifying contexts requiring the use of (case-marked) definite articles and evaluating the development of the production of such articles throughout the observation period of two years. Following procedures from previous corpus-based language acquisition research (e.g., Clahsen et al. 1994; Eisenbeiß 2003; Eisenbeiß et al. 2006), I opted for a context-based approach, hence the analysis was twofold. First, every context requiring the use of a definite article was identified and coded for the required case (nominative, genitive, dative, accusative), gender (masculine, feminine, neuter), and number (singular, plural). Second, it was coded for whether the child had produced an article, and if so, which article. The context-oriented approach enabled the visualization of the quantitative results in the form of tables (see Tables 4 and 5). This approach was especially helpful when trying to gain a first impression of the data. Nevertheless, the following sketch will focus on qualitative analyses providing descriptions of what the children are saying, since the amount of the data analyzed is insufficient for an in-depth quantitative investigation (see Defina et al. 2023b).

For the analyses, only definite articles in determiner function were considered. Many German definite articles which are used as adnominal determiners (e.g., <u>das</u> Kind schläft 'the child is sleeping') have a pronominal equivalent (e.g., schau dir <u>das</u> an 'look at that'). Pronouns, which are also marked for case, were excluded from the analyses because the referents of the pronouns often could not be identified. When a referent is unclear, the

pronoun uttered cannot be analyzed in terms of its context since the required gender and number are not identifiable. In many instances, the required case might be identifiable, but this does not suffice as there are very few one-to-one relationships between grammatical cases and articles. To ensure identifiability of the required gender and number, the article must occur with a noun, hence the focus on determiners.

I only included utterances that consisted solely of a noun or solely of a definite article and a noun when it was possible to identify the required case unambiguously, for example in answers to questions (e.g., mother: 'Which room is this?', child: '(the) kitchen'). In most other instances, the required case (as well as gender and number) of definite articles produced was clearly identifiable and it was therefore unproblematic to code the required contexts. There were only very few exceptions, for example when (case assigning) verbs were missing. These instances were not considered in the quantitative analysis but will be described and discussed as part of the sketch. In the case of article omissions, only instances with contexts requiring definite articles were considered. All instances of article omission were excluded from analysis if the given context was insufficient to determine whether a definite or an indefinite article would have been required. If nouns were inaudible, the corresponding definite articles were also excluded from analysis. If there were inaudible sequences immediately preceding a noun, these sequences were likewise excluded from analysis. All instances of exact repetition (self-repetitions and repetitions of adult utterances were not considered as part of the quantitative analysis, since including repetitions - especially of errors (e.g., von *den[ACC] Mann[DAT], von *den[ACC] Mann[DAT], von *den[ACC] Mann[DAT] 'from the man') – would distort the interpretation of the data.

In the case of prepositional phrases, articles in contractions (e.g., $zu + dem \rightarrow zum$ 'to the') were considered felicitous and analyzed like all other articles. The reduction of the articles das to s, dem to m, and den to n after prepositions in utterances like $\ddot{u}ber\underline{s}$ $Gel\ddot{a}nder$ 'over the railing' [Leo at 4;0], $auf\underline{m}$ Ball 'on the ball' [Pauline at 2;6], $auf\underline{n}$ Teller 'on the plate' [Pauline at 4;0] are infelicitous in Standard German, but were nevertheless considered felicitous for the purpose of the present analysis as they are common in colloquial German (see Cieschinger 2016 for comprehensive information on contractions of prepositions and definite articles in German). It appears reasonable to include such forms when analyzing German child language since they frequently occur in (spoken) German (Augustin 2014: 15; Cieschinger 2016), which suggests that children are likely to be familiar with these forms.

3.2 Case marking in child language

This chapter describes the marking of case on definite articles, both in NPs and in PPs², in child language. Both "non-target use of definite articles" (Bittner 2006: ch. 4) and target use of definite articles are addressed. The subsequent sections cover the acquisition steps and acquisition phenomena which could be identified in the corpus data, some of which stem from the quantitative analysis of grammatical contexts shown in Tables 4 and 5.

² NPs can of course be embedded in a PP. Here, however, for contrasting purposes, I will speak of NPs when the article and the noun are not embedded in a PP. I will speak of PPs, on the other hand, when a preposition governs the case of an article and a noun.

Table 4. Grammatical contexts (gender, number, case) of obligatory definite articles and definite articles produced by Pauline (2;0-4;0).

			lasculin singula		Feminine, singular		Neuter, singular			All genders, plural			
		Nom	DAT	Acc	Nom	DAT	Acc	Nom	DAT	Acc	Nom	DAT	Acc
	Ø	0	0	0	2	1	0	0	2	1	0	0	0
	der	0	0	0	0	0	0	0	0	0	0	0	0
. 7	die	0	0	0	0	0	0	0	0	0	0	0	0
2;0 (n = 7)	das	0	0	0	0	0	0	0	0	0	0	0	0
2;0	den	0	0	0	0	0	0	0	0	0	0	0	0
	dem	0	0	0	0	0	0	0	1	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0
	Ø	0	0	1	0	0	0	0	0	0	0	1	0
	der	5	0	0	0	4	0	0	0	0	0	0	0
2;6 (n = 25)	die	0	0	0	1	0	0	0	0	0	0	0	2
<i>(u =</i>	das	0	0	0	0	0	0	0	0	1	0	0	0
2;6 (den	0	1	1	0	0	0	0	0	0	0	1	0
	dem	0	5	0	0	0	0	0	1	0	0	0	0
	des	0	0	0	0	0	0	1	0	0	0	0	0
	Ø	0	0	0	0	0	0	2	0	0	0	0	0
	der	2	0	0	0	2	0	0	0	0	0	0	0
: 20)	die	0	0	0	3	0	1	0	0	0	1	0	1
3.0 (n = 20)	das	0	0	0	0	0	0	2	0	0	0	0	0
3;0	den	0	0	1	0	0	0	0	0	0	0	0	0
	dem	0	2	0	0	0	0	0	2	0	0	0	0
	des	0	0	0	0	0	0	1	0	0	0	0	0
	Ø	0	0	1	0	0	0	0	0	0	0	0	0
	der	0	0	0	0	8	0	0	0	0	0	0	0
: 21)	die	0	0	0	1	0	5	0	0	0	0	0	0
= u)	das	0	0	0	0	0	0	1	0	1	0	0	0
3;6	den	0	0	2	0	0	0	0	0	0	0	0	0
	dem	0	0	0	0	0	0	0	2	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0
	Ø	1	0	0	0	0	0	0	0	0	1	0	0
	der	4	0	0	0	3	0	0	0	0	0	0	0
= 35	die	0	0	0	4	0	1	0	0	0	1	0	1
$4;0 \ (n=35)$	das	0	0	0	0	0	0	2	0	3	0	0	0
4;0	den	0	0	7	0	0	0	0	0	0	0	0	0
	dem	0	5	0	0	0	0	0	2	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0

Note. \emptyset = Article omission. Cells shaded in grey indicate the required definite article in the corresponding grammatical context. The number of contexts is given in brackets. The genitive case is not included because no genitive contexts could be identified.

Table 5. Grammatical contexts (gender, number, case) of obligatory definite articles and definite articles produced by Leo (2;0-4;0).

			lasculin singula		Feminine, singular		Neuter, singular			All genders, plural			
		Nom	DAT	Acc	Nom	DAT	Acc	Nom	DAT	Acc	Nom	DAT	Acc
	Ø	0	0	0	0	0	0	0	0	0	0	0	0
	der	0	0	0	0	0	0	0	0	0	0	0	0
<u>(</u>)	die	0	0	0	0	0	0	0	0	0	0	0	0
$2;0 \ (n=0)$	das	0	0	0	0	0	0	0	0	0	0	0	0
2;0	den	0	0	0	0	0	0	0	0	0	0	0	0
	dem	0	0	0	0	0	0	0	0	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0
	Ø	0	0	0	2	0	0	1	1	0	0	0	0
	der	11	0	0	0	5 0 2 0 0 0 0 3 3 0 0 0 0 4	0	0					
52)	die	1	0	0	7	2	3	0	0	0	4	0	2
<i>(u</i> =	das	0	0	0	0	0	0	4	0	1	0	0	0
2;6 (<i>n</i> =	den	0	0	2	0	0	0	0	0	0	0	0	0
	dem	0	3	0	0	0	0	0	1	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0
	Ø	0	0	1	0	0	0	1	0	0	0	0	0
_	der	2	0	0	0	0	0	0	0	0	0	0	0
3.0 (n = 17)	die	0	0	0	2	0	1	0	0	0	0	0	0
<i>u</i>)	das	0	0	0	0	0	0	4	0	2	0	0	0
3;0	den	0	0	3	0	0	0	0	1	0	0	0	0
	dem	0	0	0	0	0	0	0	0	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0
	Ø	0	0	0	0	0	0	0	2	0	0	0	0
_	der	4	0	0	0	4	0	0	0	0	0	0	0
: 45)	die	0	0	0	14	0	2	0	0	0	5	0	1
=u	das	0	0	0	0	0	0	1	0	2	0	0	0
3;6	den	0	2	1	0	0	0	0	1	0	0	1	0
	dem	0	3	0	0	0	0	0	1	0	0	0	0
	des	1	0	0	0	0	0	0	0	0	0	0	0
	Ø	1	1	2	0	0	0	0	0	0	0	0	0
	der	2	0	0	0	2	0	0	0	0	0	0	0
: 35)	die	0	0	0	1	0	0	0	0	0	2	0	0
$4;0 \ (n=35)$	das	0	0	0	0	0	0	1	0	4	0	0	0
4;0	den	0	3	1	0	0	0	0	3	0	0	0	0
	dem	0	2	0	0	0	0	0	0	0	0	0	0
	des	0	0	0	0	0	0	0	0	0	0	0	0

Note. \emptyset = Article omission. Cells shaded in grey indicate the required definite article in the corresponding grammatical context. The number of contexts is given in brackets. The genitive case is not included because no genitive contexts could be identified.

3.2.1 Article omission

Articles, by definition, must appear with a noun (Meibauer et al. 2015: 25, 132; Becker 2021: 42, 56), and thus are not to be expected until the two-word stage. This suggests that the number of words within utterances may be a central factor in the development of case marking in German.

At the age of 2;0, Leo appears to be on the cusp of moving from the late phase of the one-word stage to the early phase of the multi-word stage, with most utterances still consisting of only one word. It should be emphasized that an analysis of mean length of utterance (MLU) was not performed as part of the present sketch, yet the vast majority of utterances in the analyzed section of the transcript (for details, see Appendix) are evidently one-word utterances. This assessment, which was merely based on 30 minutes of a single one-hour session, could be confirmed by an MLU analysis of the Leo corpus provided by Behrens (2006). According to Behrens (2006: 9), Leo began uttering word combinations at the age of 1;11. His MLU is, however, around 1.2 words at the age of 2;0 (see Behrens 2006: 12), indicating that the majority of Leo's utterances consist of one word.

As mentioned previously, utterances consisting solely of a noun could not be considered for further case analysis and were thus not classified as article omissions. There are no analyzable case contexts in the selected 30-minute section for Leo at age 2;0. Also, he does not produce any definite articles except for one article-like form: *de.* This form, which does not exist in adult German, appeared once across all ten analyzed sections of the two child language corpora. Leo (2;0) uses this potential proto-form (Peters & Menn 1993) as a substitution for a proper article (in (2)).

(2) *de Autos* 'the cars'

The correct article is presumably *die*, though this remains uncertain because of the lack of syntactic as well as conversational context. The required case can therefore not be determined. As *de* is not part of the regular paradigm of German definite articles, it does not bear any information on case or gender. Thus, *de* is case and gender neutral. It appears that this proto-form might represent Leo's first steps into the acquisition of articles as it is the only word resembling a definite article at the age of 2;0 before articles occur frequently at the age of 2;6. There was no evidence of Pauline using this or another proto-article in the analyzed data. However, Leo's data at age 2;0 might indicate that before the six German definite articles *der*, *die*, *das*, *den*, *dem*, and *des* are used, a reduced article form may occur.

In longer utterances, presumably during the early phases of the multi-word stage, definite articles appear to be omitted frequently. As can be seen in Tables 4 and 5, both children omit obligatory definite articles in every case context (except the genitive; see Section 3.2.2). Although instances of article omission occur in all of the five analyzed sections of the Rigol corpus and in four of the five analyzed sections of the Leo corpus, Pauline's session at the age of 2;0 differs from all other sessions in that Pauline omits almost every obligatory definite article (see Table 4). Beyond the instances listed in Table 4 at age 2;0, there were further instances requiring a definite article. Their (case) context is, however, not evident due to the absence of case-assigning elements (e.g., examples (4), (7) and (9)) and could therefore not be specified in the tabular format. Nonetheless, obligatory definite articles are omitted in all of these utterances except for one (see (9)) and examples of these instances will thus be given below.

- (3) *3 Tatze domm leich
 ART.DEF.F.SG.NOM⁴ cat come soon
 Pauline (2;0): '(the) cat is coming soon'
- (4) * Mäh da sea there Pauline (2;0): 'there (is the) sea'
- (5) * Oma⁵ reingehn
 ART.DEF.F.SG.NOM granny go.inside
 Pauline (2;0): '(the) granny is going inside'
- (6) * Buch einpack
 ART.DEF.N.SG.ACC book pack
 Pauline (2;0): '(I) pack (the) book'
- (7) * Auto lafen hab ich car sleep have I Pauline (2;0): 'I slept (in the) car'

Examples (3) to (7) all reflect instances of article omission. In (3), Pauline (2;0) omits the obligatory article in the nominative NP (die) Katze 'the cat'. As the gender of the noun Katze (realized as ['tatsə]) is feminine, the required article is *die.* Likewise, she omits the article in (4), yet it remains unclear what Pauline (2;0) intends to say. The investigator assumes that Pauline wants to say, "Da ist das Meer." ('There is the sea.'), indicating the absence of another nominative article. Examples (5) and (6) are comparable in that they both consist of only a noun and a verb. However, while utterance (5) lacks only the determiner of the clause's subject, utterance (6) lacks the subject itself (the context suggests the pronoun ich 'I' as intended subject) and the object's determiner. Example (7) likewise requires not only the production of an article as part of an NP, but instead requires both the production of a preposition and a determiner as part of a PP. In Standard German, the required preposition in together with the required dative article dem would be realized as a contraction (im). Another example, example (8), shows that Pauline does construct PPs at the age of 2;0. The article *dem* produced as part of the contraction *am* correctly marks the dative as required by the preposition an. Yet, the preposition an is incorrect in this context; it should be auf instead. Auf and dem could be realized without a contraction. Either way, the dative article dem would be obligatory, and therefore Pauline (2;0) produces the correct definite article, albeit an

³ The utterances quoted here and in the following examples throughout the present paper are provided as transcribed by the original authors of the corpora. Since this paper focuses on definite articles, only relevant non-target realizations with respect to the focus of interest are marked with an asterisk. The children's utterances might also be deviant regarding phonological features, word order, or other linguistic features. However, these deviations are not marked with an asterisk.

⁴ In the case of erroneous productions of definite articles (marked with an asterisk), the correct (i.e., the required) form is glossed. Glossing an incorrect form would require arbitrary decisions, since no German definite article is fully unambiguous regarding its marking of gender, number, and/or case (see Table 3). Detailed interlinear glossing is also provided for omitted articles, provided the syntactic context is clear.

⁵ Kinship names are often used as proper nouns, that is, without an article. However, there are also contexts where article use is obligatory. In this situation, Pauline was looking out the window and commenting on someone she saw outside. Shortly before (5) she said, ein Oma 'a grandma'. It was clear from the investigator's utterances that this was not the child's grandmother, suggesting the need for an article. In general, I treated definite articles as non-obligatory for kinship names. Utterances like (i.a) Mama macht Abendbro(t) and (i.b) die Mama macht Abendbrot 'mom is making dinner' [Leo at 2;6] were treated as equally correct. I excluded utterances like (i.a) from the analysis due to the non-obligatory context and thus did not classify these instances as article omissions. However, since utterances like (i.b) are examples of definite article use, I included such instances in the analysis.

incorrect preposition. In example (9), Pauline (2;0) produces a definite article, but she omits the (case assigning) verb or preposition. It is therefore uncertain which case should be marked on the missing article. Based on immediately preceding utterances containing the verbs *passen* 'to fit' and *reingehen* 'to go in', the NP *den Auto* 'the car' could be interpreted as a subject ("Das Auto passt da rein."; N.SG.NOM). In any case, the article *den* is certainly incorrect as it not part of the neuter paradigm (see Table 3).

- (8) da sit am Tlo there sit at.ART.DEF.N.SG.DAT loo Pauline (2;0): '(it) is sitting in the litter box'
- (9) *den Aute da rein
 car there in
 Pauline (2;0): 'the car (fits) in there'

At the age of 2;0, Pauline produces only two definite articles within the analyzed session: one as part of a PP (in (8)) and the other in an NP (in (9)), which is erroneous. It may be concluded that Pauline (2;0) is yet to discover the role of definite articles, which might explain why she omits most obligatory definite articles (see Table 4). At the same time, these two articles might reflect that first definite articles can occur as early as the age of 2;0.

Later in acquisition, article omissions (examples in (10), (11) and (14)) are rarely found in Pauline's utterances. At the ages 2;6, 3;0, 3;6, and 4;0, only one or two articles are omitted in 20 to 35 obligatory contexts (see Table 4). Therefore, one may conclude that between the ages of 2;0 and 2;6, Pauline discovers the role of definite articles in language production in German and begins to use these articles in grammatically required contexts. Omissions of obligatory definite articles can be found more frequently in Leo's utterances (see Table 5, examples in (12) and (13)). Yet, between the ages of 2;0 and 4;0 (see Table 5), a developmental stage similar to that described for Pauline (2;0), characterized primarily by article omissions alongside the use of very few definite articles, was not observed in Leo's sketch data.

- (10) sneidst du dir nich in * Finger? cut you you not in ART.DEF.M.SG.ACC finger Pauline (2;6): 'you don't cut your finger?'
- (11) * Nudeln wern schon leer

 ART.DEF.F.PL.NOM noodles are.becoming already empty
 Pauline (4;0): '(the) pasta is almost empty already'
- (12) und dann schwimmt und Dock hebt sich and itself then swims and ART.DEF.N.SG.NOM dock rises Leo (3;0): 'and then (it) swims and (the) dock rises'
- soll (13)aber sie nich an Dom dran but she should not on ART.DEF.M.SG.ACC cathedral thereon Sutzgebiet und nich protected.area and not on.ART.DEF.N.SG.ACC

Leo (4;0): 'but she should not touch (the) cathedral or the protected area'

3.2.2 The role of the genitive

Out of the four cases of the German language, the genitive article *des* seems to play the least important role in child language. Throughout the five hours of data analyzed, not a single genitive context could be identified, hence there was no need for the children to mark the

genitive with a definite article. Yet, throughout the ten sessions there were some occurrences of the article *des*. Pauline (2;6 and 3;0) and Leo (3;6) produce the article *des* as the determiner of a noun in an accusative (in (14)) or nominative (in (15) and (16)) context.

- (14) in * nächsten Wochen kannst du mit mir in ART.DEF.F.PL.DAT next weeks can you with me *des Puzzle machen ART.DEF.N.SG.ACC jigsaw.puzzle make
 - Pauline (2;6): 'in (the) next few weeks you can do the puzzle with me'
- (15) des is *des Bad this is ART.DEF.N.SG.NOM bathroom Pauline (3;0): 'this is the bathroom'
- (16) guck das is *des Korkenzieher look that is ART.DEF.M.SG.NOM corkscrew Leo (3;6): 'look, that's the corkscrew'

Pauline only uses *des* in neuter contexts, replacing the required and phonologically similar article *das*, possibly reflecting a dialectal issue.⁶ There are three indicators suggesting this. Firstly, in the transcripts '[: das]' is written after every 'des', indicating that Pauline intends to say 'das' but produces an assimilation instead. Such an annotation is otherwise not provided when Pauline produces a non-target article, such as *den* instead of *dem* (in (17)). Secondly, although Pauline commonly produces the neuter article *das* correctly, she generally produces the demonstrative pronoun *das* (this, that) as *des* (see example (15)). Thirdly, when listening to the audio recordings of the sessions, Pauline's mother can often be heard saying [dɛs] instead of /das/ both in pronominal and in determining function, although the transcribers always transcribed these instances as 'das'.

Contrary to Pauline (in (14) and (15)), Leo (in (16)) used the article *des* instead of the required article *der* not *das*. Further analysis of this instance is, however, not useful. Due to the nature of the data, the fact that there is only a single instance of productive use of *des*, and the fact that the parental input was not analyzed, hypothesizing about Leo's knowledge of and his exposure to the article *des* would be inappropriate.

3.2.3 Overgeneralization of accusative to dative contexts

The context-oriented approach chosen for the present analysis allowed comparison of the use of the various articles in all case, gender, and number contexts. Throughout the data, a pattern became apparent: accusative articles seem to be prone to overgeneralization to dative contexts. More precisely, the masculine singular accusative article *den* appears to be frequently overgeneralized to *dem*-contexts.

As can be seen in Table 3, the definite article *dem* marks masculine and neuter singular dative contexts. Besides *des*, *dem* is the only unique definite article in that it solely marks one case. The article *den*, which bears a strong phonological similarity to the article *dem*, serves as a marker for masculine singular accusative and plural dative contexts. There were several

⁶ The realization of *das* as [dεs], written either <des> or <däs>, has been described for various dialects of German, for example, South Hessian (Weiß & Dirani 2019), the dialect of Heidelberg (Klein & Rieck 1982) and Styrian vernaculars (Maierhofer 2017). However, since the region where data collection took place is unknown, it cannot be verified whether this phenomenon also occurs in the local dialect.

instances throughout the analyzed sections of the corpora where the children used the definite article *den* in a non-target manner.

The following sections on non-target-like occurrences of *den* will be structured based on gender of the noun. The nouns in the phrases in question (see (9), (17)–(27)) show no signs of plural inflection. Therefore, all contexts will be treated as singular contexts.

Non-target-like occurrence of den in masculine singular contexts

Both Pauline and Leo produce the article den in contexts other than accusative contexts. While Pauline only produces one such utterance (in (17)), Leo produces several utterances of this kind (in (18)–(22)).

- (17) der liest *den Papa nen Buch vor this.one reads_17 ART.DEF.M.SG.DAT daddy a book ptc_1 Pauline (2;6): 'this one reads a book to his daddy'
- (18) weil auf *den Hinweg die Sonne mich because on ART.DEF.M.SG.DAT way.there ART.DEF.F.SG.NOM sun me geblendet hat dazzled has
 - Leo (3;6): 'because the sun dazzled me on the way there'
- (19) kennst du die Kassette mit *den blauen Mond? know you ART.DEF.F.SG.ACC cassette with ART.DEF.M.SG.DAT blue moon Leo (3;6): 'do you know the cassette with the blue moon?'
- (20) weil sie an *den Hampelmann zieht because she on ART.DEF.M.SG.DAT jumping.jack pulls Leo (4;0): 'because she is pulling the jumping jack'
- (21) da wischte man mit *den Fuß drüber there wiped one with ART.DEF.M.SG.DAT foot over Leo (4;0): 'you are to wipe your foot on that'
- (22) halt Wilhelmine auf *den Soß bitte hold Wilhelmine on ART.DEF.M.SG.DAT lap please Leo (4;0): 'hold Wilhelmine on your lap, please'

The examples (17) to (22) show that non-target-like use of *den* occurred only in contexts where *dem* and therefore the dative would be required. Several occasions of an overgeneralization of *den* to masculine singular dative contexts (in (18)–(22)) could be identified throughout the analyzed sections of the Leo corpus. However, every erroneous use of *den* of this kind produced by Leo occurs in the context of a PP. Two of the produced prepositions, *auf* and *an*, are two-way prepositions. It may be assumed that two-way prepositions pose a greater challenge in acquisition and might be prone to overgeneralizations to one case (i.e., accusative to dative). The preposition *mit*, however, exclusively assigns the dative. Yet the article Leo (3;6) produced in (19) is erroneous, nonetheless. Occurrences of *den* as a determiner of a masculine noun without a preceding preposition (like in (17)) could not be observed in Leo's utterances.

⁷ Separable verbs, such as the verb *vorlesen*, are composed of a prefix (e.g., *vor*-) and a verb (e.g., *lesen* 'to read'). In some grammatical constructions, like in (17), the prefix must be separated from the verb (Eisenberg 2020a: ch. 7.1.2/3). To indicate that these elements belong together, the interlinear translations of the verb and the corresponding verb particle are marked with an underscore and a matching number.

Non-target-like occurrence of den in neuter singular contexts

Like non-target-like occurrences of *den* in masculine singular contexts, *den* also occurs frequently in neuter singular contexts. Likewise, Leo produces more such utterances (in (23)–(27)) compared to Pauline (in (9)).

- (23) da die ganze Farbe ab-gefallen an *den Schiff there ART.DEF.F.SG.NOM entire paint off-fallen on ART.DEF.N.SG.DAT ship Leo (3;0): 'all the paint has flaked off the ship there'
- (24) Thorsti hat sich mit den Händen an *den
 Thorsti has himself with ART.DEF.F.PL.DAT hands on ART.DEF.N.SG.DAT

 Gras abgestützt
 grass supported
 - Leo (3;6): 'Thorsti supported himself with his hands on the grass'
- (25) vielleicht is dann alles zu bei *den Hochbett maybe is then all closed by ART.DEF.N.SG.DAT bunk.bed Leo (4;0): 'maybe the bunk bed (would be) closed off (securely)'
- (26) das wächst auf *den Feld that grows on ART.DEF.N.SG.DAT field Leo (4;0): 'that grows in the fields'
- (27) smeckt das *den Swein gut? tastes that ART.DEF.N.SG.DAT pig well Leo (4;0): 'is the pig enjoying that?'

For all of Leo's utterances of this kind ((23)–(27)) it can be observed again that the article *den* in neuter singular contexts occurred only in dative contexts. Again, most errors occur within a PP (in (23)–(26)). While the same two-way prepositions, *an* and *auf*, are again produced with the same error pattern as in masculine contexts, there is once more an instance of a preposition governing solely one case (i.e., the dative): *bei*. Example (27) shows the only case of an overgeneralization of the article *den* to a dative *dem*-context outside a PP.

Pauline (2;0) produces the article *den* in a non-target way in only one utterance with a neuter singular context (in (9)). This utterance also marks the only instance of the occurrence of *den* in a (potential) nominative context (see Section 3.2.1 for discussion) across the five hours of data analyzed.

No occurrence of den in feminine singular contexts

The definite article *den* (as well as the article *dem*) cannot be used in feminine singular contexts in Standard German. In line with this, there were no instances of an occurrence of *den* (nor of *dem*) in feminine singular contexts across the five hours of analyzed data.

At the same time, overgeneralizations of accusative to dative contexts are difficult to detect, since both accusative and nominative feminine nouns require the article *die*. Therefore, it would be inadequate to attribute the occurrence of *die* in dative contexts (as in (28) and (29)) to an overgeneralization of the accusative – it could also be attributed to an overgeneralization of the nominative.

(28) mit *die Mama with ART.DEF.F.SG.DAT mom Leo (2;6): 'with mom' (29) so mit *die Hand like.this with ART.DEF.F.SG.DAT hand Leo (2:6): 'like this with the hand'

Overall, this type of error appears to be rather negligible as the two utterances in (28) and (29) are the only examples of this phenomenon across the analyzed sections of the ten sessions. Furthermore, both utterances were produced by Leo during one session (2;6) and the erroneous use of *die* only occurred after the preposition *mit* in both instances. Moreover, Leo (2;6) seems to be unsure about which article to use in example (29), as the inconsistent use of the definite article suggests. In example (31), an utterance which Leo (2;6) produced shortly before (29), he uses the correct dative article *der* in an otherwise identical utterance.

Summary

Taken together, we can make several assumptions. The article *den* is prone to overgeneralization to *dem*-contexts. When used in a non-target manner, *den* occurred in singular dative contexts in all instances but one (in (9); but note the uncertainty about the required case in this utterance) and was never uttered in plural contexts. While in plural contexts the article *den* can also mark the dative, none of the nouns in the phrases in question show any signs of plural inflection. Therefore, it seems more plausible to attribute the deviations to errors of case due to overgeneralization of the accusative than to assume that the incorrect use of the article *den* reflects an error of number.

Based on the analyzed data, the overgeneralization of the accusative article den seems to occur mainly in PPs. This can at least be stated for Leo, while Pauline makes hardly any errors involving the overgeneralization of the definite article den. A difference between prepositions governing only one case and two-way prepositions did not become apparent. The examples provided in (18) to (26) could suggest that Leo uses all prepositions with accusative case. However, there are also several examples at the ages 2;6, 3;6, and 4;0 that suggest otherwise (samples given in (30)–(33)).

- (30) die is am Bahnhof Oschatz this.one is at.ART.DEF.M.SG.DAT station Oschatz Leo (2;6): 'this one is at Oschatz station'
- (31) nur mit der Hand only with ART.DEF.F.SG.DAT hand Leo (2;6): 'only with the hand'
- (32) die U_Bahn is auf der Hebebühne
 ART.DEF.F.SG.NOM subway is on ART.DEF.F.SG.DAT lifting.platform
 Leo (3;6): 'the subway is on the lifting platform'
- (33) im Kindergarten haben wir ein Fußlappen in.ART.DEF.M.SG.DAT kindergarten have we a foot.rag Leo (4;0): 'in kindergarten we have a foot rag'

To conclude, it seems that throughout the course of acquisition, overgeneralizations of accusative to dative can occur – especially in masculine and neuter singular contexts. The opposite phenomenon, the overgeneralization of dative to accusative (especially the overgeneralization of *dem*- to *den*-contexts) could not be observed in the analyzed sections of the two corpora.

3.2.4 Summary

Based on the analysis of five hours of spontaneous speech data from Leo and Pauline (2:0-4:0), four hypotheses can be made about the acquisition of case marking and definite articles in German. First, it can be conjectured that a reduced article form, a proto-article, may occur before the first appearance of full definite article forms. However, because this hypothesis is based on a single observation, it should be treated with appropriate caution. Second, children seem to omit definite articles frequently during the early multi-word phase. This pattern was particularly evident in one of Pauline's samples. Third, it is hypothesized that the genitive case plays the least important role among the four cases in German child language. Fourth and last, based on repeated observations, it is hypothesized that the article *den* is frequently overgeneralized to dative contexts.

4 Comparison with acquisition literature

The example sketch presented in this paper demonstrates that the analysis of five hours of natural conversation data from two children across the ages 2;0 to 4;0 as suggested by Defina et al. (2023a) can reveal error patterns and developmental steps in the acquisition of a language. In order to further explore the novel sketch approach, the results of the provided sketch of the acquisition of German, a well-studied language, will be compared to results from existing acquisition literature on German.

For this purpose, findings from ten studies will be consulted, namely Clahsen (1984), Mills (1985), Clahsen et al. (1994), Tracy (1986), Eisenbeiß (2003), Bewer (2004), Bittner (2006), Eisenbeiß et al. (2006), Scherger (2018), and Szagun (2019). These studies are diverse in terms of their research focus, the number of children studied, the developmental stages investigated (i.e., the ages of the children), as well as the nature of the data used (e.g., longitudinal vs. diary studies). Clahsen (1984) reported on developmental stages in monolingual German children's acquisition of case. He studied three children longitudinally: two boys between 1;5 and 3;5 (fraternal twins) and their younger sister (1;1 to 2;4). In 1985, Mills provided an extensive report on the acquisition of German in the first volume of Slobin's handbook series The Crosslinguistic Study of Language Acquisition (Slobin 1985). In her report, Mills (1985) summarized and analyzed data from children of different ages from a variety of sources, focusing on diary studies (e.g., Preyer 1882; Scupin & Scupin 1928). Tracy (1986) analyzed longitudinal language data from six children. Four children were aged 1;6 to 3;0, one child was aged 1;0 to 3;0, and the sixth child was aged 1;0 to 4;10. Clahsen et al. (1994), Bewer (2004) and Bittner (2006) used the data of Simone (1;9 to 4;0 years) from the Miller corpus (Miller 1979), an extensive longitudinal corpus of three girls acquiring German monolingually in the 1970s, as their database. Clahsen et al. (1994) investigated the acquisition of case marking and its relation to the development of phrase structure. Bewer (2004) studied the acquisition of articles as gender markers and Bittner (2006) focused on the acquisition of case- and genderrelated features of adnominal definite articles and corresponding pronominals. Eisenbeiß (2003) and Eisenbeiß et al. (2006) analyzed case marking in cross-sectional and longitudinal corpora of spontaneous and semi-spontaneous speech of seven monolingual German children aged 1;11 to 3;6 and five children between 2;6 and 3;6. Note that these studies used partially the same data (see Eisenbeiß 2003: 197 and Eisenbeiß et al. 2006: 13). Scherger (2018) investigated dative marking in spontaneous and semi-spontaneous speech of nine monolingual German children (mean age: 4;9) amongst further groups of children (i.e., older, bilingual and/or with a developmental language disorder). Given the special focus on dative constructions, I decided to use this study for comparison with regard to dative phenomena in the sketch, despite the slight deviation in age. Szagun (2019) provided an analysis of 16

children between the ages of 1;4 and 2;10 and another six children between the ages of 1;4 and 3;8. The children were frequently recorded across the investigation period. The purpose of the data collection was to yield a comprehensive dataset of child language as well as child directed speech in German and to function as control data for data from children with cochlear implants (Szagun 2004). Overall, the subject of case acquisition in German has been covered in a wide range of studies, which will serve as reference for the following comparison.

4.1 Acquisition phenomena found in the sketch

The analysis of five sections each of the Leo corpus and the Rigol corpus revealed frequent omission of definite articles during the early phases of the multi-word stage. This phenomenon was primarily observed in Pauline at the age of 2;0, when she omits nearly every obligatory definite article. Six months later, both Pauline and Leo omit hardly any definite articles. These results are in line with Mills (1985: 174), who describes that definite articles start to occur at the age of 2;0 but are frequently omitted until the age of 2;6. According to Tracy (1986: 50), children start marking case overtly once their utterances are longer than three or four words. Clahsen (1984: 8) also notes that children in the early stages of language acquisition, up to the age of 2;10, produce primarily NPs without determiners. Such a phase of frequent article omission could not be found for Leo among the five investigation points at 2;0, 2;6, 3;0, 3;6, and 4;0. At the age of 2;6 and subsequent time points, Leo produces definite articles in the majority of required instances. Before that, at the age of 2;0, Leo has presumably just reached the early multi-word stage.

At this age, Leo (2;0) produces one article-like form: de (in (2)). This is the only instance of a child producing such a proto-form in the ten analyzed sections of the present paper. However, this form stands out as it is the only form resembling a definite article at the age of 2;0. The acquisition literature shows that such a case and gender neutral proto-form appears to occur frequently. Tracy (1986: 62, 73) found that children frequently use the protoarticle de instead of a standard definite article, which is then gradually replaced by standard forms over the course of acquisition. Bittner (2006: 121) furthermore states that children's use of de for definite articles reflects the functional reduction to their [+definite] feature. In Szagun's study with typically developing children, there is also evidence of the proto-form de (2019: 108). Mills (1985: 174) and Bewer (2004: 114) likewise describe the occurrence of de as an early definite article form. In sum, the occurrence of de appears to be a frequent phenomenon before children start producing full definite articles. Leo might have just started using this form at the age of 2,0 and there might be further instances of the occurrence of de in other sessions, that were not analyzed for the provided sketch. Due to the long gaps between the sampling points, the likelihood to miss critical periods for phenomena with short duration is high. The sketch data suggest that this proto-form exists, but the occurrences are too rare to investigate this acquisition process.

Once Leo and Pauline have stopped omitting articles and begin using case-marked definite article forms, articles occur for all case forms except one from age 2;6. Throughout the five hours of data, there is not a single instance of a grammatical context requiring a genitive article. However, there are three instances in which the genitive article *des* occurs (in (14)–(16)). According to Kauschke (2012: 78), empirical studies investigating case acquisition in German often do not even take the genitive into account, since it is generally accepted that the genitive is negligible as an objective case in German child language. When investigating the role of the genitive in German language acquisition, Mills (1985: 155) found that children do not mark the genitive on articles up to the age of 4;0. Marking of the genitive

solely occurred in the form of prenominal genitives (Mills 1985: 185). Ruff (2000) and Eisenbeiß et al. (2009) report similar observations in their studies on the acquisition of possessive structures. Thus, the previously described absence of the genitive on the basis of five hours of acquisition data does not appear to be an artefact of the sketch format, but rather seems to represent the minor role the genitive plays amongst the four German cases in child language (but see Ulrich & Motsch 2018, for details on productive correctness in genitive contexts in children aged 4;0 to 8;11).

In contrast, there are numerous instances of utterances in the sketch data that contain the accusative article den. Den is frequently overgeneralized to dative contexts. This phenomenon was identified both in masculine and neuter singular contexts, with most instances occurring in PPs. These results largely correspond to those of Clahsen (1984), Mills (1985), Tracy (1986), Eisenbeiß (2003), Bittner (2006), Eisenbeiß et al. (2006) and Scherger (2018), which are summarized in Table 6. Across the data analyzed for the provided sketch, no instances of overgeneralizations from *dem* to *den* contexts were found. It should be noted, however, that such errors may occur in relation to certain word orders (Eisenbeiß 1994) or animacy patterns (Drenhaus & Féry 2008). In both Leo's and Pauline's data, correctly marked dative articles appeared as early as the age of 2;6. Pauline even produces a correct dative form at age 2;0 (in (8)), which appears to be earlier than in many other children (see Table 6). Pauline generally makes very few mistakes in dative contexts. In a few instances, she omits the article altogether (e.g., in (14)) and only in one instance, in (17), she substitutes dem with den, although this single error does not occur after a preposition, which the literature suggests is a common source of error. However, when Leo produces erroneous articles in dative contexts, most errors occur in PPs and all these errors are overgeneralizations of the accusative. In contrast, Pauline's sketch data does not exhibit this pattern. She makes hardly any overgeneralization errors. Rather, from the age of 2;6, she produces most accusatives and dative forms correctly, even after prepositions. Overall, the findings in the literature (see Table 6) indicate that the overgeneralization of the accusative to dative contexts described in the previous sketch, especially in PPs, is a typical phenomenon in the acquisition of German. Mills (1985: 155), Szagun (2019: 109-110) and Eisenbeiß et al. (2006: 11) attribute this confusion to the high phonological similarity of the articles den and dem. Another explanation may be that children have not yet acquired the lexeme-specific case-assignment properties of certain verbs or prepositions (see Eisenbeiß 1991, summarized in Eisenbeiß 2003: 308).

Mills (1985)

Phenomenon	Reference
Den often occurs in dative contexts.	Bittner (2006)
Dem is scarcely ever overgeneralized to other grammatical cases.	Bittner (2006)
Accusative forms are frequently overgeneralized to dative contexts – especially after prepositions that govern the dative.	Clahsen (1984), Mills (1985), Tracy (1986), Eisenbeiß (2003), Eisenbeiß et al. (2006) Scherger (2018)
Overgeneralizations from dative to accusative contexts occur never or only rarely.	Clahsen (1984), Mills (1985), Scherger (2018)
Dative markings appear from the second half of the third year of life.	Tracy (1986)
Dative markings first occur around age 3;0.	Mills (1985)
The dative is generally marked correctly as soon as its	Mills (1985)

Table 6. Phenomena described in the literature related to the acquisition and overgeneralization of accusative and dative.

4.2 Acquisition phenomena not found in the sketch

first forms occur, except in PPs.

There are also phenomena which are described in the literature but are not attested by the sketch analysis: 1) the occurrence of the first definite articles; 2) frequent overgeneralizations of the nominative to other cases; and 3) the temporary presence of a case system consisting of only the nominative and the accusative.

According to Bewer (2004: 114) and Mills (1985: 174), the first definite articles occur around age 2;0. The snapshot nature of the sketch does not allow for the precise dating of the emergence of first forms. Leo produces no definite articles during the first session, but many are present six months later. This suggests that Leo's first full article forms occur between 2;0 and 2;6. Pauline already produces some definite articles at 2;0, suggesting that first forms may have occurred earlier.

The first marking of case on articles appears, according to Mills (1985: 178), in the nominative. In addition, Mills observes that the nominative is frequently overgeneralized to accusative contexts during the early stages of case acquisition (1985: 178). Both Clahsen (1984: 9) and Clahsen et al. (1994: 105) identify the same pattern, adding that the nominative also occurs in dative contexts. In the data analyzed for the sketch above, the nominative articles *der* and *das* never occur in other case contexts besides nominative contexts. Solely the article *die* falsely occurs in two feminine singular dative contexts (in (28) and (29)), although it is unclear whether these represent an overgeneralization of the nominative or the accusative (see Table 3). It remains uncertain whether Pauline and/or Leo underwent a phase with frequent overgeneralization of nominative forms. Ultimately, however, the sketch data of this study is not suited to resolve this matter. For Leo, such a phase might have taken place between the investigation points at 2;0, where he produces no articles except for a one-time proto-form, and 2;6, where he produces definite articles in nominative, accusative, and dative contexts. At the first investigation point at 2;0, Pauline omits nearly every definite article, though there is one utterance containing a dative article in the correct context (in (8))

and another instance containing the article *den* (in (9)), yet the required case is unclear (for discussion see Section 3.2.1).

Furthermore, some studies have identified a stage in which children exhibit a "binary case system" (Clahsen 1984: 21; Tracy 1986: 54; Clahsen et al. 1994). During this stage, only the nominative and the accusative are being used, with the accusative occurring in both accusative and dative contexts. Such a stage could not be identified in the sketch data. The limited amount of sketch data may explain the absence of this pattern in the sketch.

5 Discussion

Taken together, the present study has four main results. 1) There were no wrong predictions. All phenomena identified in the sketch data are valid. 2) Despite large individual differences, the sketch data revealed indications of developmental trajectories. 3) The sketch does not capture all phenomena described in the acquisition literature. Some developmental stages take place in specific, sometimes narrow, time windows, while the sampling method has large gaps of six months each. Therefore, the sketch data may not comprise all relevant data in the corpus. 4) Beyond this, infrequent phenomena (e.g., the genitive in German child language) are especially unlikely to occur in the sketch corpus.

Although the database of the sketch is limited to only five hours, all main findings could be confirmed by the existing (case) acquisition literature. Nevertheless, some acquisition phenomena were identified in only one of the two children, most likely due to individual variation in language acquisition, supporting the approach taken by Defina et al. (2023a), which recommends recording a minimum of two children.

At the same time, the sketch grammar model has limitations. The 30 minutes of each session only allow for a snapshot of what a child is saying at a certain age. In addition, there are large gaps in between sampling points that risk missing relevant phenomena in acquisition. Furthermore, compared to elicited language, spontaneous language always carries the risk of over-representing individual forms and under-representing other forms. Like most of the acquisition studies which were used for comparison, the sketch was based solely on spontaneous language. The sketch's database was, however, very limited as opposed to the database of previous acquisition studies. Looking at the structured analyses performed on the basis of Pauline's and Leo's utterances, we can see that some forms did not occur at certain ages. There are no plural contexts throughout the 30-minute sections of the sessions of Pauline at the ages 2;0 and 3;6 (see Table 4) and of Leo at age 3;0 (see Table 5). Likewise, there are no masculine singular contexts in Pauline's data at the age of 2;0. The absence of these contexts in the tabular format can arise both due to the absence of these forms across all utterances (i.e., the child does not produce such contexts, for example, the child does not produce any plural nouns) or due to a lack of analyzability (e.g., omission of case-assigning verbs in the context of masculine nouns). On the other hand, at age 3;6, both Pauline (see Table 4) and Leo (see Table 5) produce a disproportionately large number of feminine singular contexts (i.e., 14 out of 21 analyzed contexts for Pauline), with Leo producing predominantly nominative forms (i.e., 20 out of 45 analyzed contexts are feminine singular contexts; 14 of these are nominative forms). Such disproportionate distributions can never be completely ruled out in spontaneous language data. However, the larger the amount of data, the more likely it is that the proportions of relevant contexts will be evenly distributed.

The sketch of German case acquisition provided in the present paper allowed for the generation of hypotheses which could be confirmed by previous studies. These previous

studies differed in their approach from the sketch approach described here. On the one hand, they had a larger database, but on the other hand, most studies examined not only definite articles, but also pronouns and other case-marked elements. Beyond this, some studies included quantitative analyses while others focused solely on qualitative descriptions. Mills (1985) for example, who studied the acquisition of case mainly based on data from diary studies, chose a similar approach to the approach taken in the sketch provided, in that she described frequent error patterns supported by a variety of examples from the data.

There were various options to structure the sketch of child language. The first option was to describe the two children, Pauline and Leo, separately and to outline what they say at each investigation point. A second option was to structure the sketch according to the five age levels (2:0, 2:6, 3:0, 3:6, 4:0), including descriptions of what both children say at these levels. Thirdly, there was the option of choosing a structure based on the four German cases, going through each case, and providing examples of target- and non-target-like marking of the respective case. The fourth option was to consider what both children say in any given case context, possibly independent of age, and to try to identify (error) patterns. I pursued this option because it enabled a broad perspective and allowed for detecting possible individual variation in language acquisition. Since the sketch format is based on a small amount of data, it cannot be expected to reveal all language acquisition phenomena of a given language. Nevertheless, the sketch format was expected to have the potential to uncover the most evident processes. It therefore seemed sensible to adopt a process-oriented approach from the outset, rather than focusing on individual characteristics or differences by strictly separating the descriptions of the children, age groups or cases. Ultimately, even with the structural approach chosen, individual characteristics and differences between the children could still be detected and described.

Overall, based on the provided sketch, the novel sketch approach to language acquisition research suggested by Defina et al. (2023a, 2023b) appears to be a suitable method to reduce the typological gap in this field.

6 Summary and conclusion

The present paper put the novel sketch grammar model proposed by Defina et al. (2023a, 2023b) to the test. To evaluate whether this approach offers a new way of gaining insight into and gathering information about first language acquisition in minority languages, I presented an exploratory sketch of the acquisition of a well-studied language. The acquisition of case marking on definite articles in German was analyzed and described based on ten 30-minute sections of two children at the ages 2;0, 2;6, 3;0, 3;6, and 4;0. This led to a comparatively small database of only five hours. While some phenomena described in existing studies of German case acquisition drawn from larger databases could not be identified in the sketch, all main findings of the sketch analysis could be confirmed by these existing studies. This suggests that the sketch format cannot be expected to reveal all language acquisition phenomena of a given language but does have the potential to uncover the most evident acquisition processes.

The sketch provided above demonstrated that the approach suggested in the SAM does indeed yield valid sketch data. As the approach in this paper was restricted to one linguistic phenomenon, further comparable studies are certainly needed to test the possibilities and limitation of the sketch format thoroughly. Fortunately, large corpora of child language and acquisition studies exist for various languages, making this a feasible undertaking. Based on the experiences with this sketch of German child language, I would like to encourage this approach.

Acknowledgements

I presented the first ideas for this paper at the 2nd Workshop on the Acquisition of Lesser-studied Languages at the University of Melbourne in 2019 and I would like to thank the participants for their comments and interesting questions. I am grateful to Heike Behrens who made the data from both corpora available to me. Many thanks to Barb Kelly and an anonymous reviewer for their critical and helpful feedback on this paper. I thank Wanyima Wighton for proofreading an earlier version and Katherine Revius for proofreading the near-final version. Finally, I want to thank Birgit Hellwig for supervision and constructive feedback from the initial idea to the final paper.

Appendix Session names, times analyzed, and location of recording.

		,		C	
	Age	Session name ^a	Time analyzed (hh:mm:ss)	Number of utterances of the focal child ^d	Location
	2;0	Leo/020006	00:33:00 - 01:03:00	372	At home
	2;6	Leo/020612	00:00:30 - 00:32:42	334	At home
0	3;0	Leo/021129b	00:02:15 - 00:32:15	260	At home
Leo	3;6	Leo/030622	00:00:00 - 00:30:00	208	At home
	4;0	Leo/040018	00:23:30 - 00:53:30	260	At home (children's room)
	2;0	Pauline/020005	00:00:00 - 00:30:00	216	INV's lounge
	2;6	Pauline/020613	00:00:00 - 00:27:48 (entire session)	270	INV's lounge, at home (children's room) ^e
ne	3;0	Pauline/030011	00:05:00 - 00:35:00	282	INV's lounge
Pauline	3;6	Pauline/030612	00:00:00 - 00:30:00	272	At home (children's room)
	4;0	Pauline/040007	00:00:00 - 00:12:15 00:17:45 - 00:29:00 00:33:30 - 00:37:00 00:39:30 - 00:42:30°	240	INV's lounge

Note. INV = Investigator; aSession name as found on CHILDES (MacWhinney, 2000); bThe file is called 021129, although the recording took place on Leo's third birthday; cThis session was divided into multiple sections to yield 30 minutes of data. This allowed for the exclusion of times when Pauline was either not near the microphone or not very talkative; dThe number of utterances given here corresponds to the number of segments in the original transcript within the analyzed sections; cPauline did not want to be recorded at INV's place, which is why the recording was continued at Pauline's place.

References

Augustin, Hagen. 2014. Zum Beispiel im Deutschen: Wenn Präposition und Artikel verschmelzen: Eine Korpusuntersuchung zu auf + definitem Artikel [For example in

- German: When preposition and article merge: A corpus study of auf + definite article]. *IDS Sprachreport* 30(3). 14–21.
- Becker, Laura. 2021. *Articles in the world's languages*. (Linguistische Arbeiten 577). Berlin & Boston: De Gruyter.
- Behrens, Heike. 2006. The input-output relationship in first language acquisition. *Language and Cognitive Processes* 21(1-3). 2-24.
- Bewer, Franziska. 2004. Der Erwerb des Artikels als Genus-Anzeiger im deutschen Erstspracherwerb [The acquisition of the article as a gender marker in German first language acquisition]. *ZAS Papers in Linguistics* (33). 87–140.
- Bittner, Dagmar. 2006. Case before gender in the acquisition of German. *Folia Linguistica* 40(1-2). 115–134.
- Brugman, Hennie & Albert Russel. 2004. Annotating Multi-media / Multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal.
- Cieschinger, Maria. 2016. The contraction of preposition and definite article in German: Semantic and pragmatic constraints. PhD Thesis, Osnabrück University.
- Clahsen, Harald. 1984. Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache [The acquisition of case markings in German child language]. *Lingustische Berichte* (89). 1–31.
- Clahsen, Harald, Sonja Eisenbeiß & Anne Vainikka. 1994. The seeds of the structure: *A syntactic analysis of the acquisition of case marking*. In Teun Hoekstra & Bonnie D. Schwartz (eds.), *Language acquisition studies in generative grammar* (Language acquisition & language disorders 8), 85–118. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Defina, Rebecca, Shanley E. M. Allen, Lucinda Davidson, Birgit Hellwig, Barbara F. Kelly & Evan Kidd. 2023a. Sketch Acquisition Manual (SAM), Part II: The sketch corpus. In Birgit Hellwig, Shanley E. M. Allen, Lucinda Davidson, Rebecca Defina, Barbara F. Kelly & Evan Kidd (eds), *The acquisition sketch project*, 5–38. (Language Documentation & Conservation Special Publication 28.) Honolulu: University of Hawai'i Press.
- Defina, Rebecca, Shanley E. M. Allen, Lucinda Davidson, Birgit Hellwig, Barbara F. Kelly & Evan Kidd. 2023b. Sketch Acquisition Manual (SAM), Part II: The acquisition sketch. In Birgit Hellwig, Shanley E. M. Allen, Lucinda Davidson, Rebecca Defina, Barbara F. Kelly & Evan Kidd (eds), *The acquisition sketch project*, 39–86. (Language Documentation & Conservation Special Publication 28.) Honolulu: University of Hawai'i Press.
- Drenhaus, Heiner & Caroline Féry. 2008. Animacy and child grammar: An OT account. *Lingua* 118(2). 222–244.
- Eisenbeiß, Sonja. 1991. Passiv und Kasus Evidenz aus der deutschen Kindersprache [Passives and case Evidence from German child language]. State examination paper, University of Cologne.
- Eisenbeiß, Sonja. 1994. Kasus und Wortstellungsvariation im deutschen Mittelfeld. Theoretische Überlegungen und Untersuchungen zum Erstspracherwerb [Case and word order variation in the German midfield. Theoretical considerations and studies on first language acquisition]. In Brigitta Haftka (ed.), Was determiniert Wortstellungsvariation?: Studien zu einem Interaktionsfeld von Grammatik, Pragmatik und Sprachtypologie [What determines word order variation? Studies on an interaction field of grammar, pragmatics and language typology], 277–298. Opladen: Westdeutscher Verlag.

- Eisenbeiß, Sonja. 2003. *Merkmalsgesteuerter Grammatikerwerb: Eine Untersuchung zum Erwerb der Struktur und Flexion von Nominalphrasen* [Feature-driven grammar acquisition: An investigation of the acquisition of structure and inflection of noun phrases]. PhD Thesis, Heinrich Heine University Düsseldorf.
- Eisenbeiß, Sonja, Susanne Bartke & Harald Clahsen. 2006. Structural and lexical case in child German: Evidence from language-impaired and typically developing children. *Language Acquisition* 13(1). 3–32.
- Eisenbeiß, Sonja, Ayumi Matsuo & Ingrid Sonnenstuhl. 2009. Learning to encode possession. In William McGregor (ed.), *The expression of possession* (The expression of cognitive categories (ECC) 2), 143–212. Berlin: Mouton de Gruyter.
- Eisenberg, Peter. 2020a. *Grundriss der deutschen Grammatik. Das Wort* [Outline of German grammar. The word] 5th edn. Berlin: J.B. Metzler.
- Eisenberg, Peter. 2020b. *Grundriss der deutschen Grammatik. Der Satz* [Outline of German grammar. The sentence], 5th edn. Berlin: J.B. Metzler.
- *ELAN* (5.9) [Computer software]. 2020. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from: https://archive.mpi.nl/tla/elan.
- Gutzmann, Daniel & Katharina Turgay. 2011. Funktionale Kategorien in der PP und deren Zweitspracherwerb [Functional categories in the PP and their second language acquisition]. Zeitschrift für Sprachwissenschaft 30(2). 169–221.
- Hentschel, Elke & Harald Weydt. 2021. *Handbuch der deutschen Grammatik* [Handbook of German Grammar], 5th edn. Berlin & Boston: De Gruyter.
- Kauschke, Christina. 2012. *Kindlicher Spracherwerb im Deutschen. Verläufe, Forschungsmethoden, Erklärungsansätze* [German language acquisition in children. Developmental trajectories, research methods, explanatory approaches] (Germanistische Arbeitshefte 45). Berlin: De Gruyter.
- Kidd, Evan & Rowena Garcia. 2022. How diverse is child language acquisition research? *First Language* 42(6). 703–735.
- Klein, Wolfgang & Bert-Olaf Rieck. 1982. Der Erwerb der Personalpronomina im ungesteuerten Spracherwerb [The acquisition of personal pronouns in unguided language acquisition]. Zeitschrift für Literaturwissenschaft und Linguistik 10(45). 35–70.
- Lieven, Elena & Sabine Stoll. 2013. Early communicative development in two cultures: A comparison of the communicative environments of children from two cultures. *Human Development* 56(3). 178–206.
- MacWhinney, Brian. 2000. *The CHILDES project. Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Maierhofer, Katja. 2017. *Tendenzen und Ebenen in der Dialektverschriftung am Beispiel steirischer Mundarten* [Tendencies and levels in dialect transcription using the example of Styrian vernaculars]. MA Thesis, University of Vienna.
- Meibauer, Jörg, Ulrike Demske, Jochen Geilfuß-Wolfgang, Jürgen Pafel, Karl Heinz Ramers, Monika Rothweiler & Markus Steinbach. 2015. *Einführung in die germanistische Linguistik* [Introduction to German linguistics], 3rd edn. Stuttgart & Weimar: J.B. Metzler.
- Miller, Max (ed.). 1979. The logic of language development in early childhood. Berlin & Heidelberg: Springer-Verlag.
- Mills, Anne E. 1985. The acquisition of German. In Dan Isaac Slobin (ed.), *The crosslinguistic study of language acquisition, vol. 1: The data*, 141–254. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Peters, Ann M. & Lise Menn. 1993. False starts and filler syllables: Ways to learn grammatical morphemes. *Language* 69(4). 742–777.
- Preyer, Wilhelm. 1882. Die Seele des Kindes [The child's soul]. Leipzig: Grieben.
- Ruff, Claudia. 2000. Besitz und Besitzer in der Sprachentwicklung von deutschen und italienischen Kindern [Possession and possessor in language development of German and Italian children]. PhD Thesis, Technische Universität Carolo-Wilhelmina zu Braunschweig.
- Scherger, Anna-Lena. 2018. German dative case marking in monolingual and simultaneous bilingual children with and without SLI. *Journal of Communication Disorders* 75. 87–101.
- Scupin, Clara & Wilhelm Scupin. 1928. *Die Kindersprache. Eine psychologische und sprachtheoretische Untersuchung* [Child language. A psychological and language-theoretical investigation]. Leipzig: Barth.
- Slobin, Dan Isaac (ed.). 1985. *The crosslinguistic study of language acquisition, vol. 1: The data.* Hillsdale, NJ: Lawrence Erlbaum Associates.
- Szagun, Gisela. 2019. *Sprachentwicklung beim Kind* [Language development in the child], 7th edn. Weinheim: Beltz.
- Tracy, Rosemarie. 1986. The acquisition of case morphology in German. *Linguistics* 24(1). 47–78.
- Ulrich, Tanja & Hans-Joachim Motsch. 2018. Die Existenz des Genitivs in der deutschen Kindersprache [The existence of the genitive in German child language]. *Forschung Sprache* 6(2). 43–55.
- Ulrich, Tanja, Martina Penke, Margit Berg, Ulrike Lüdtke & Hans-Joachim Motsch. 2016. Der Dativerwerb Forschungsergebnisse und ihre therapeutischen Konsequenzen [Dative acquisition Research findings and their therapeutic consequences]. *Logos* 24(3). 176–190.
- Weiß, Helmut & Seyna Maria Dirani. 2019. Strong or weak? Or: How information structure governs morphosyntactic variation. In Antje Dammel & Oliver Schallert (eds.), *Morphological variation: Theoretical and empirical perspectives* (Studies in Language Companion Series 207), 311–342. Amsterdam: John Benjamins Publishing Company.

Received: 1 Aug 2022 Accepted: 6 Mar 2023