Journal of Child Language (2024), **51,** 1450–1477
doi:10.1017/S0305000923000521

ARTICLE

Processing of noun plural marking in German-speaking children: an eye-tracking study

Bénédicte GRANDON[1], Marcel SCHLECHTWEG[2] and Esther RUIGENDIJK[1]

[1]Institute of Dutch Studies, University of Oldenburg and Cluster of Excellence "Hearing4all"
[2]Institute for English and American Studies, University of Oldenburg and Cluster of Excellence "Hearing4all"
Corresponding author: Bénédicte Grandon; Email: Benedicte.Grandon@uni-oldenburg.de

(Received 28 July 2022; revised 05 May 2023; accepted 30 August 2023)

Abstract

The ability to process plural marking of nouns is acquired early: at a very young age, children are able to understand if a noun represents one item or more than one. However, little is known about how the segmental characteristics of plural marking are used in this process. Using eye-tracking, we aim at understanding how five to twelve-year old children use the phonetic, phonological, and morphological information available to process noun plural marking in German (i.e., a very complex system) compared to adults. We expected differences with stem vowels, stem-final consonants or different suffixes, alone or in combination, reflecting different processing of their segmental information. Our results show that for plural processing: 1) a suffix is the most helpful cue, an umlaut the least helpful, and voicing does not play a role; 2) one cue can be sufficient and 3) school-age children have not reached adult-like processing of plural marking.

Keywords: noun plural marking; L1 Acquisition; inflection processing; eye-tracking

Introduction

Language acquisition studies have been exploring children's abilities to use linguistic information for comprehension and production. Inflection in general and plural marking in particular have been given a lot of attention over the years. This is especially true for production of plural marking but less so for its comprehension and processing. Moreover, noun plural marking has been mainly considered through the prism of morpho-phonology, leaving out the role of phonetic information on its processing and comprehension. Our study investigates the comprehension and processing of plural marking in German, a language with a very complex plural marking system. We use eye-tracking to understand how phonetic, phonological and morphological information in plural marking are processed and acquired by children. We are especially interested in understanding how the segmental characteristics of the material used in noun plural marking are processed: are all of these characteristics used to a similar extent? We argue that the various cues are processed differently due to their location in the inflected form (i.e., part of the stem or suffix), the type of mechanism they result from (i.e., a phonological, morphological process or a subsequent phonetic adaptation) and their combination (i.e., are they provided alone or in combination with each other).

Acquisition of plural marking in children

Children understand the conceptual difference between singular and plural using available morphosyntactic information, as early as two years of age (e.g., J. N. Wood et al., 2009). This early competence translates in expressive morphosyntax, as they start producing noun plural marking around the age of two. This ability develops in a so-called U-shaped curve as has been shown in studies with English-speaking children: they inflect non-words using the correct phonologically-conditioned affix (-s, -z, or -ɪz) at three years of age, as seen in the well-known study by Berko (1958), and also produce both regular and irregular plural forms of nouns (Marcus, 1995). Following this early period, they then start to make errors: most notably, errors of overgeneralization (i.e., using a regular inflection instead of an irregular one), as shown in the studies of Marcus (1995) on English, and Behrens (2002, 2011) on German. Around the age of six, their production evolves towards more adult-like productions, with fewer errors in both regular and irregular plural forms. Starting from these errors and overgeneralizations made by children, Clahsen (1999) proposes a "dual-route" model for processing inflection: irregular forms are stored in the lexicon and retrieved through lexical processes, whereas regular forms are retrieved by means of decomposition (in reception) and recomposition (in production), following grammatical rules.

While English relies exclusively on overt suffixation (with a few lexicalized exceptions), other languages use both stem changes and suffixation. Several studies explored how these types of plural marking are acquired in production. Ravid and Schiff (2009) suggested that plural marking in six- and seven-year-old Hebrew-speaking children improves over time, with better performance with non-changing stems over changing stems, and regular over irregular suffixes. Albirini (2015) showed that in Jordanian Arabic, by the age of eight years, all predicable patterns are acquired but not all other, less predictable patterns, with productivity and frequency playing a role in early stages, and predictability in later stages of acquisition. Saiegh-Haddad et al. (2012) further indicated that plural marking is still not adult-like at seven or eight years in Palestinian Arabic. In this study, accuracy of plural marking was higher when it involved suffixation rather than stem changes.

Few studies have examined how well children comprehend plural marking. A series of studies showed that, for English, young children were able to understand plural marking at the age of two years when it is marked with an /s/ suffix but not with the /z/ allomorph, (Davies et al., 2017), and were not fully able to understand the syllabic allomorph /ɪz/ as plural marking until three years of age (Davies et al., 2020). After this age, the three allomorphs of English were well understood by the children, yet the representations of singular and plural were shown to develop gradually, and at different rates. This was evidenced by a lower accuracy in the identification of singular items when compared to plural items (Davies et al., 2019). Taken together, these studies of plural comprehension and processing in English indicated effects of the type of allomorphs (segmental vs. syllabic) and a possible role of their acoustic salience in plural comprehension and acquisition.

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Using a judgment task, the only study on plural comprehension in German-speaking children showed that sensitivity to well-formedness of plural forms was increasing in German-speaking children between the age of six to ten years, however without being adult-like (Korecky-Kröll et al., 2012).

These studies provide an overall outline of early processing abilities of plural marking in children, yet a number of questions remain unanswered, especially regarding the links between the forms of the plural marking and their processing and acquisition by children.

Our choice to study processing in noun plural marking in German-speaking children is motivated by the diversity of mechanisms and material available in German to mark the plural. It especially allows to manipulate different types of cues used alone or in combination with each other to mark the noun plural, and explore how the characteristics of these different types of cues influence the processing of plural marking.

Noun plural marking in German

German has a complex noun plural marking system, relying on different plural markers, resulting from different morpho-phonological processes, and subsequent phonetic adaptation processes. Köpcke (1988) distinguishes eight different morpho-phonological formation processes for German, applying to all or certain genders (i.e., masculine, feminine and neuter) and relying on suffixation and vowel change, alone or in combination. Four different suffixes are available: segmental suffix /s/, syllabic suffixes /ə/ and /ɐ/, and /n/, which can be both segmental as [n] or syllabic as [ən]. Another type is formed by "umlaut", a vowel fronting of the back vowels /u/ and /o/, and the low-vowel /a/, resulting in /y/, /ø/ and /ɛ/, respectively. It can occur as the sole plural marker or in combination with the syllabic suffixes /ə/ and /ɐ/. Finally, a noun's singular and plural forms can also be identical: disambiguation of the noun number will then only be possible through the determiner, as in das_{SG} Messer [mɛsɐ] – die_{PL} Messer [mɛsɐ], *the knife - the knives.* Table 1 below, based on Köpcke (1988) summarizes all possible plural markings per gender.

As shown in several examples in Table 1, plural is not only marked by phono-morphological material, but also results in different phonetic changes located on the word's stem-final consonant. For example, in the Kind_{SG}-Kinder_{PL} pair, the stem-final consonant is the location of a voicing [d]-[t] alternation, following German's word-final devoicing rule (i.e., /d/ realized as [t] in word-final position); in the Tochter_{SG}-Töchter_{PL} pair, there is an alternation of the fricatives [x] and [ç] consecutive to the fronting of the stem's vowel (i.e., [ɔ] changed to [œ]).

Two factors should be taken into account in the description of German plural marking: Type frequency and productivity. Using a corpus analysis based on the 1000 most frequent lemmas in the DeReWo corpus (Institut für Deutsche Sprache, 2009), Zaretsky et al. (2013) indicate that not all plural markers are encountered equally frequently in German. These plural markers are different for each gender (see Table 2): -(e)n is the most frequent plural type overall and for feminine nouns in particular, -e is the most frequent for neuter nouns, and -e (with or without umlaut) or no suffix are the most frequent types for masculine nouns.

Productivity is the second relevant factor to describe plural type in German: different markers are used more than others to inflect new words in the language and are thus more productive. Productivity is therefore important in the study of how plural marking is being acquired or processed. As indicated in Laaha et al. (2006), the -s suffix is the most

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Table 1. Plural marking in German (adapted from Köpcke, 1988; with IPA transcriptions of standard German pronunciation from Krech et al., 2009)

	Gender		
	masculine	feminine	Neuter
Def. article (sing./plur.)	der/die	die/die	das/die
Plural morpheme			
-e, [ə]	Fisch-Fische	Kenntnis-Kentnisse	Jahr-Jahre
	[fiʃ]-[fiʃə]	[kɛntnis]-[kɛntnisə]	[ja:ʶ]-[ja:ʁə]
	fish	*knowledge*	*year*
-(e)n, [ən][n]	Bauer-Bauern	Tür-Türen	Auge-Augen
	[baɔ̯ɐ]-[baɔ̯ɐn]	[ty:ᵄ]-[ty:ʁən]	[aɔ̯gə]-[aɔ̯gən]
	farmer	*door*	*eye*
-er, [ɐ]	Geist-Geister	–	Kind-Kinder
	[gaɛ̯st]-[gaɛ̯stɐ]		[kɪnt]-[kɪndɐ]
	ghost		*child*
-s, [s]	Park-Parks	Mutti-Muttis	Auto-Autos
	[paʶk]-[paʶkə]	[mʊti:]-[mʊti:s]	[aɔ̯to:]-[aɔ̯to:s]
	park	*mom*	*auto*
ø	Adler-Adler	–	Fenster-Fenster
	[a:dl̩ɐ]-[a:dl̩ɐ]		[fɛnstɐ]-[fɛnstɐ]
	eagle		*window*
umlaut	Bruder-Brüder	Tochter-Töchter	Kloster-Klöster
	[bʁu:dɐ]-[bʁy:dɐ]	[tɔxtɐ]-[tœçtɐ]	[klo:stɐ]-[klø:stɐ]
	brother	*daughter*	*monastery*
umlaut + -e	Sohn-Söhne	Kuh-Kühe	Floß-Flöße
	[zo:n]-[zø:nə]	[ku:]-[ky:ə]	[flo:s]-[flø:sə]
	son	*cow*	*raft*
umlaut + -er	Wald-Wälder	–	Volk-Völker
	[valt]-[vɛldɐ]		[fɔlk]-[fœlkɐ]
	wood		*people*

productive plural type in German, as it can be used with all three genders, in loanwords and with little phonological constraints, whereas umlaut or -er suffix + umlaut plural types are not productive at all (i.e., they are never applied to new words).

The dual-route model discussed by Clahsen (1999) stresses the role of another factor in processing and acquisition of plural markers: regularity, which according to Clahsen (1999) is the property for plural types to be formed following a strict rule-based mechanism, or as proposed by Crystal (2008) the property to "be in conformity with

Table 2. Frequency of plural types in German, adapted from Zaretsky, et al. (2013).

		Gender		
Suffix	All genders	masculine	feminine	Neuter
-s	6%	5%	3%	15%
-e	22%	32%	1%	47%
-(e)n	43%	11%	92%	4%
-er	4%	2%	0%	17%
-e + umlaut	12%	24%	4%	1%
umlaut	1%	0%	0%	0%
no suffix	13%	25%	0%	17%

the general rules of a language". Even though they are not the most frequent plural types, -s suffixes are considered to be the most regular and therefore the default plural markers in German. They are indeed the best candidates to be processed using rule-like decomposition mechanisms. In contrast, in this approach, plural words formed with plural markers considered to be less regular, such as -n or -e are processed as individual lexical entries. Note that the status of -s as the only regular and possible default plural marker has been debated, for example, by Penke and Krause (2002) who argued that -s was not the only regular plural type in German, but that the -n suffix for feminine nouns is regular as well.

These characteristics of noun plural marking in German (number of allomorphs, type frequency, productivity and regularity) already give us an indication of the complexity of its system, that can potentially be difficult to acquire, both in production and in comprehension. The cues to plural marking are defined on three levels – namely, morphology (i.e., suffixation), phonology (i.e., stem vowel changes or umlaut) and phonetics (i.e., stem-final consonant voicing alternation). Differences in productivity are seen in morphological materials (i.e., suffixes) over phonological processes (i.e., umlaut) and between the different types of suffixes (e.g., -s is more productive than other suffixes). Frequency analyses are further indicating a preference for the use of suffixation alone over the combined use of suffixation and stem vowel change and finally for the use of stem vowel change alone, as shown on Table 2 above. Strikingly, analyses of productivity and frequency are not including the phonetic level (i.e., voicing alternation of the stem-final consonant), even though it might provide an additional cue signaling a plural form.

Another characteristic could be relevant to plural marking, both in its processing and acquisition – namely, the salience of the plural markers. This is related to the salience of their segmental information, which we define as the segments' property to stand out in the speech stream or within the phonological system. Phonetically, salience can be understood as the perceptual correlate of the sound's acoustic properties (e.g., "perceived resonance of major classes of speech sounds", Clements, 2009) and phonologically, it can be seen as the result of differences in sonority of speech sounds within the phonological system (e.g., Selkirk, 1984). Following these definitions, vowels are more salient than consonants, segments in stressed positions are more salient than those in unstressed positions, voiced consonants are more salient than voiceless consonants. Finally, a syllabic

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

suffix is more salient than a segmental suffix, as it uses more material and leads to resyllabification of the stem, that is a change of the stem's prosodic structure.

Acquisition of plural marking in German-speaking children

German-speaking children are confronted with a complex plural marking system, in comparison to other languages, such as English, which only has three allomorphs. This complexity has been at the center of several studies on the production of plural marking in children. Overgeneralization or overregularization (i.e., systematic, preferred use of one type when others are expected) are the focus of several studies: in these studies, -s and -n suffix often are the overgeneralized plural types, in addition to or in substitution of other cues. In a longitudinal study from the ages of 1;1 to 4:0, Behrens (2002, 2011) showed that all plural types are already present in the inventories of German-speaking children at age two. As predicted by the U-shape development curve, these children quickly turned to overgeneralizations before evolving to a more adult-like use of plural markings. By age four, generalizations (i.e., using the -n suffix instead of other markings) still occurred, along with filling-gap strategies (i.e., adding material when the plural form is similar to the singular form). This is consistent with the results of the production study by Clahsen et al. (1992), which showed overregularizations using -n suffixation in children. Kauschke et al. (2011) further indicated that plural production improves in children until reaching a plateau around the age of five without being adult-like, even at 6;11. They further showed that the accuracy of plural marking is better for high-frequency words, and when plural is marked by -e and -n suffixes, than with -s suffix, and lower when plural is marked only with a vowel change (i.e., umlaut). Picture-elicitation tasks were used by Laaha et al. (2006) to study the effects of productivity of the plural marking on their acquisition by 2;6- to 6-year-old, German-speaking children and by Laaha et al. (2015) to study cross-linguistic differences in acquisition in 4;1- to 5;7-year-old children (with normal hearing and with cochlear implants) with German or Dutch as their L1. Their results with German-speaking children showed higher accuracy for plural types with -e suffixes (with or without vowel change) than with -er suffix with vowel change, compared to -s suffixes, -n suffixes, and finally with vowel change only, with significant effects of age, plural type, productivity. For both Dutch- and German-speaking children, accuracy was higher when plural marking was more predictable (i.e., when the choice of the plural marking is phonologically determined), more transparent (i.e., when there is no stem change) and with higher stem frequency (i.e., lemma frequency). The plural marking system is much less complicated in Dutch, and accuracy of plural marking was unsurprisingly higher in Dutch-speaking children. Finally, Zaretsky et al. (2013, 2016) showed different types of overgeneralization patterns in children who spoke German as an L1 (i.e., preference for -s suffixes) or as an L2 (i.e., preference for -n suffixes and zero forms). Children with German as an L1 further tended to use only the -s suffix to mark plural on non-words, when adults tended to use both -s and -n suffixes in the same task.

As mentioned previously, the study by Korecky-Kröll et al. (2012) is the only study in German-speaking children aged six to ten years and adults on plural marking comprehension. This study used well-formedness judgments: speakers were presented the plural form of a word orally along with a matching picture, and were asked if it was correctly produced. Stimuli were split across three conditions: the actual plural form of the noun, a morphologically illegal plural form and a potential but non-existing plural

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

form. Age effects were observed, and even older children (aged ten years) did not reach adult-like ceiling performance in their judgment of existing forms. Both token frequency (measured from the CELEX database, Baayen et al., 1993) of the plural forms and productivity of the plural marking were found to affect the children's responses. Cues such as umlaut or suffixes were considered with regards to their frequency and productivity, and how they contribute to the potentiality (i.e., acceptable in the language's morphology but not existing) or illegality (i.e., not acceptable in the language's morphology) of the plural form. However, no analyses of how these cues of the plural markers are used for the comprehension were provided. For example, there was no indication that a specific cue to plural marking such as umlaut or suffixation helps or hinders the comprehension of plural forms more than another.

In sum, the acquisition of plural in general and in German-speaking children in particular is overwhelmingly studied in production, and only a handful of studies to date have looked at the children's comprehension abilities. Production is shown to be affected in different ways by factors such as word (i.e., lemma) frequency, or predictability and productivity of the plural marking. The role of its morpho-phonological form is only studied as segmental vs. syllabic, and to our knowledge, more specific analyses of the plural marking have not been conducted, in any language.

Research questions

Our literature review showed mechanisms at play in noun plural marking processing and its acquisition by children in general and by German-speaking children in particular. Both production and comprehension have been discussed, without however considering how the phonetic and phonological information carried by plural markings are processed. Language acquisition uses oral information, naturally since a large part of this process occurs before the children start to learn how to write, which most often starts around the age of six years. Therefore, early representations at higher linguistic levels such as morphology or syntax are built on oral information provided. Studies on the acquisition of noun plural have shown that this process is influenced by factors such as regularity, frequency or productivity of the plural markers. However, to our knowledge, no study has focused on how children use the information available in the speech signal and its varying salience to acquire this specific language feature.

With this in mind, our study aims at understanding how the processing of plural noun marking is affected by the linguistic levels (i.e., phonetic, phonological and morphological), the combinations and locations (i.e., stem-medial position for umlaut, stem-final position for voicing alternation or stem-external position for suffixes) of the different cues used to mark noun plural, thus raising the following research questions:

What kind of information (phonetic, phonological, morphological) do speakers of German use to process noun plural marking? Do the type and number of cues available affect the processing and comprehension of plural differentially? Do school-age children use this information in a way similar to adults?

Answering these research questions will help us clarify the role of phonetic, phonological and morphological information on inflection processing. To this end, we use an eye-tracking paradigm, which allows for fine-grained analyses of how decision and processing evolves over time: we use accuracy of response as an indicator of the ability of the participant to identify the singular or the plural form of a given word, reaction time

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

(RT) as a measure of certainty of this identification of the singular or plural form, and gaze fixation as a measure of processing leading to decision.

We hypothesize that the phonetic, phonological and morphological information carried by the plural markers are not used in the same way. If salience is a phonetic or phonological characteristic of segments, we expect that changes in stem vowel and vocalic suffixes -/ə/ and -/ɐ/ are the most salient cues to plural, followed by voiceless sibilant -/s/ and nasal -/n/ suffixes. If salience is linked to the level of processing, we expect suffixation (i.e., morphological process) to be more salient, since it contrasts with empty material in the singular form, followed by stem vowel change (i.e., umlaut, a phonological process on a stressed vowel), and finally we expect voicing alternation of the stem-final consonant to be the least salient cue to plural marking (i.e., a phonetic adaptation following a suffixation).

Finally, since plural processing might also be related to non-linguistic factors, we will further examine if and how individual characteristics of the children, such as their age, their vocabulary development and their Phonological Working Memory abilities affect the comprehension and processing of plural marking.

Method

Participants

Thirty children and 32 adults participated in the study. Children's age ranged from 5;01-11;10 (mean: 8;03, SD: 2;00) and adults' age ranged from 18;11-46;08 (mean: 24;09, SD: 6;03). No child had a history of speech and language pathology, as reported per parental questionnaire. All participants were screened for hearing impairment by means of an audiogram: all had PTA4 scores within normal hearing levels.

All participants were native speakers of standard German spoken in Northern Germany. Several participants (adults and children) reported a second language or dialect spoken at native level: five adults reported a variety of Low German (non-standard, regional varieties); one adult reported Turkish and Azerbaijani; two children reported English; and one child reported Italian as a second native language. The children gave oral consent: a written consent was given by one legal representative for the child participants, and by the adult participants themselves. All were informed that they could end their participation at any time, without justification.

The recruitment and testing procedures were approved by the Ethical Commission of the University of Oldenburg, and followed the COVID19 guidelines for studies involving human participants in place at that time at the University of Oldenburg.

This study was part of a larger project including several language tasks, which took about 1:30-2:00 hours per participant – in one testing session (adult participants) or two testing sessions (child participants) – which was/were divided in several blocks with breaks in between. Adult participants received a small financial compensation and children a small present after each session for their participation.

Testing procedure

Adults were tested in one session in the following order: audiogram followed by the two halves of the plural processing task. Children were tested in two sessions: in session one, the audiogram was followed by a non-word repetition task, and the first half of the plural

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

processing task and after one week, session two included the second half of the plural processing task and a vocabulary development test.

Vocabulary development and phonological working memory were assessed by means of two tests: Peabody Picture Vocabulary Test (PPVT) and LITMUS – Non-word repetition test.

Peabody Picture Vocabulary Test (PPVT), 4th edition– German version

The German version of the Peabody Picture Vocabulary Test (PPVT, Lenhard et al., 2015) is a standardized test of receptive vocabulary development, aimed at children from five to 14 years. It consists of 228 items, in blocks of 12 and takes about 25 minutes to administer. In this test, an experimenter pronounces an item, and asks the child to point to the matching picture out of a four-picture set. Accuracy of the picture selection is measured and compared to normative values. All child participants to our study received scores within the normal or upper range of the test's distribution.

LITMUS-NWR – German version

The Non-Word Repetition test (NWR, Chiat, 2015; Hinnerichs, 2016 for the German version) is a non-standardized production test aiming at assessing phonological working memory abilities in children. We used the German version of the test. The test consists of 16 pre-recorded items. The items are sequences of two to five Consonant-Vowel (CV) syllables. They are presented on a computer and the child is asked to repeat each sequence of syllables. The child's production is recorded for later assessment. For our purposes, a whole-word accuracy was rated for each item. The scores of the children of our study ranged from 10 to 16 (total: 16 items).

Experimental task

We used a picture-selection task with eye-tracking to test the comprehension and processing of nominal plural in children.

Plural types and word selection

We departed from the traditional classification of plural (e.g., Köpcke, 1988; Wiese, 2009) based on gender and written form, and we instead defined and restricted our analyses to ten categories using phonetic, phonological and morphological information of plural markings. Our categories were created by manipulating three types of phonetic, phonological and morphological material alone or in combination, to mark plural: stem-medial vowel change (i.e., umlaut); suffix addition to the singular form; and subsequent intervocalic voiced realization of the stem-final stop consonant. This resulted in four categories: 1) plurals with no change compared to the singular; 2) plurals with one change which is either a stem-medial vowel change (i.e., umlaut), or a suffix; 3) plurals with two changes, which are a suffix associated with a stem-medial vowel or stem-final consonantal change; or 4) plurals with three changes which are a stem-medial vowel change, a stem-final consonantal change and a suffix. The details of these plural types and number of differences between singular and plural forms are presented in Table 3 with examples to

Table 3. plural types in our study – types and number of contrasting cue available to mark plural

Plural type	Type of contrasting cues available: Vowel change (umlaut)	Suffix	Voicing contrast	Number of cues to mark plural (suffix/umlaut/ voicing contrast)	Example (singular-plural pairs)
No change	–	–	–	0	Kissen-Kissen [kɪsn] - [kɪsn], *pillow*
Vowel change: umlaut	+	–	–	1	Vogel-Vögel [foːgl̩] – [føːgl̩], *bird*
–/s/ suffix	–	+	–		Auto-Autos [ao̯toː] - [ao̯toːs], *car*
–/ə/ suffix	–	+	–		Tisch-Tische [tɪʃ] – [tɪʃə], *table*
–/n/ suffix	–	+	–		Katze-Katzen [katsə] – [katsən], *cat*
Umlaut + -/ə/ suffix	+	+	–	2	Ball-Bälle [bal] – [bɛlə], *ball*
Umlaut + -/ɐ/ suffix	+	+	–		Schloss-Schlösser [ʃlɔs] – [ʃlœsɐ], *castle*
Voiced stem-final C + /ə/ suffix	–	+	+		Pferd-Pferde [pfeːᵆt] – [pfeːᵆdə], *horse*
Voiced stem-final C + /ɐ/ suffix	–	+	+		Kleid-Kleider [klaɛ̯t] – [klaɛ̯dɐ], *dress*
Umlaut + Voiced stem-final C + /ə/ or /ɐ/	+	+	+	3	Wald-Wälder [valt] – [vɛldɐ], *forest*

For each plural type, we chose eight different words. These words had to be depictable and understood easily by children (i.e., objects, animals, people…); they had to be singular-dominant (e.g., words like "eye" were excluded); and lemma frequency had to be as homogeneous as possible within and across categories. Lemma frequency was assessed using the Intellitext 2.6 database (Wilson et al., 2010): mean frequency (SD) across plural types was 43.40 (98.88) instances per million in four combined corpora. Details per plural type are given in Appendix 1. The total number of items was: 10 plural type * 8 words *2 numbers = 160 items. The items were presented in blocks of 20 items, balanced in number (10 plural and 10 singular targets), plural type (two of each type), randomized, and split into two testing sessions of 80 items each. Each testing session of 80 items took about 10-15 minutes to complete.

illustrate them. A complete list of items is provided in supplementary materials (https://osf.io/8t3pe/).

Material

All audio stimuli were recorded by a female native speaker of standard German spoken in the Northwest of Germany. The recording took place in a sound-attenuated booth at the University. Words were presented in singular-plural pairs on a computer screen. Recordings were made with a Neumann KM 184 cardioid Microphon, a Gina Echo 3G audio interface and the Adobe Audition software (sampling frequency 44.1 kHz, mono,

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

32 bits, wav format). The audio files were then cut and scaled in intensity (60dB SPL) using Praat (Boersma & Weenink, 2020) to ensure constant volume of item presentation for the duration of the study. For each stimulus, the time point when each item of a singular-plural pair starts to differ from the other member of the pair (e.g., the onset of the stem vowel in /fuːs/-/fyːsə/ *<foot>)* was extracted with a Praat script, for data analyses.

For the picture stimuli, we used black-and-white drawings of objects, animals or people retrieved from the Pixabay (https://pixabay.com) and Shutterstock (https://www.shutterstock.com) databases. We used one item for pictures depicting the singular form of the word, and two copies of the exact same item (same drawing, same size) for pictures of the corresponding plural form of the word. The position on the screen (left-right) for singular and plurals was balanced, 50% singular on the left side and 50% on the right side.

Procedure and data recording

Before starting each 80-item testing session, a 5-point calibration was conducted, to ensure appropriate detection of the participant's eye-gaze.

A blank slide was presented for 1500 ms before each item. We then first presented a fixation cross in the middle of the screen on a white background. After 500 ms of cross fixation, two pictures depicting the singular and the plural version of the item were presented. After 500 ms, the pre-recorded audio target word was played through two speakers placed on each side of the monitor and set at a comfortable level. After hearing the word (singular or plural version), participants were asked to select the picture matching that word in pushing the left or right button on a Chronos response box. They were instructed to answer as soon as they could after the word ended. Before the testing phase, a training phase with six items was run, to ensure understanding of the task and familiarity with the response device by the participants. Figure 1 illustrates the visual and audio presentation of the stimuli.

A Tobii Pro Fusion eyetracker was used with a laptop and an external screen (size: 21.5", 476.57mm x 268.76mm), on a one-computer setting. We ran the experiment and collected behavioral data with Eprime 3.0 (Psychology Software Tools, Inc. [E-Prime 3.0], 2016). We recorded eye movements with the Eprime Extensions for Tobii (TET Package), with a 250 Hz sampling frequency.

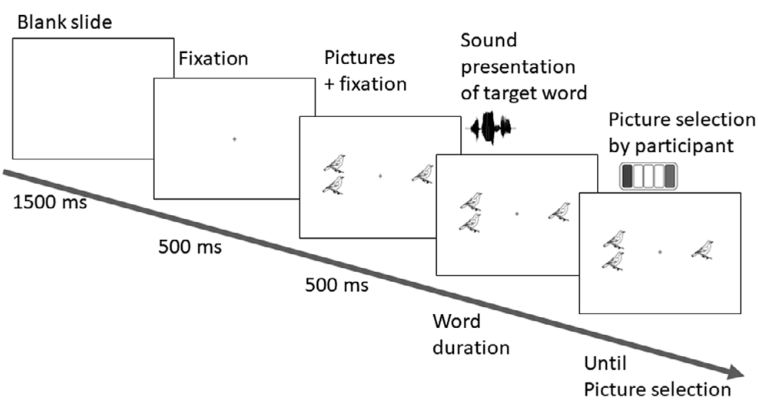

Figure 1. Stimuli presentation procedure.

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Data analyses

Data preprocessing

Accuracy is measured as the correct or incorrect identification of the target item presented orally. RT corresponds to the time between the word offset and the participant's response, and log-transformed to reduce the influence of potential outliers.

For the gaze fixation analyses, we first calculated the number of fixations towards two areas of interest (AOI) – namely, the target picture and the competitor – in 100 ms timebins, centered on the item's first singular-plural formal divergence. We excluded fixations to the other areas of the screen from this count. We then calculated a percentage of fixations towards the target (vs. competitor) for each 100 ms timebin.

Items for the "no change" category were used as a "baseline" condition, and allow us to study a possible bias towards singular or plural forms. However, they were then excluded from further analyses, since they do not provide differences between singular and plural forms: no accuracy can be computed and gaze data could not be aligned to a specific time point.

The TET Package of the Eprime extensions for Tobii recorded movements of both eyes every 4 ms, providing a validity assessment (1 for valid measurement, 0 for no valid measurement) at each timepoint. We computed a percentage of valid measurement per item and participant (for a total of 25 per 100ms-timebin). No participants had to be excluded; however, 86 items (out of the 8856 total items presented, i.e., < 1% of the total) had to be excluded because of insufficient measurements (they had less than 75% valid data points).

Statistical modeling. Accuracy: GLMM

The dependent variable for the analyses of Accuracy is a binomial (0/1) response. Therefore, it was analyzed with a Generalized Linear Mixed-Effect Model (GLMM). We used the *glmer()* function from the *lme4* package (Bates et al., 2015) with the R software, version 4.0.3 (R Core Team, 2020). A first model was created, and one by one, all possible independent variables were added: Group, Number, and either 1) Plural Type (9 levels, see Table 3 above) or 2) Number of differences between singular and plural forms. Model improvement was assessed through step-by-step model comparisons based on AIC scores (Akaike Information Criterion, Akaike, 1974). A first, simple model is built. We then add a single variable of interest to this first model to obtain a second, more complex model. We then compare the two models' AIC scores: if the AIC is significantly lower in the second, more complex model, we assume that the added variable improves the fit of the model; however, if the AIC of the second, more complex model is not significantly lower, we assume that the variable of interest is not improving the model's fit, and this variable is rejected; we then consider a new variable to be added; this process is repeated for each variable of interest, alone and in interaction with variables already included in the model. Categorical predictors such as Group (Adults vs. Children), Number (Plural vs. Singular), Plural types (9 levels) and Number of cues (1, 2 or 3) were included in the GLMM using a deviation coding, which allows each level of a predictor to be compared to the mean accuracy of all other levels.

Statistical modeling. Reaction time: LMM

The dependent variable of our analyses of log-transformed Reaction Time (RT) is a continuous variable and was thus analyzed by means of a Linear Mixed-Effect Model (LMM) using the *lmer()* function from the *lme4* package (Bates *et al.*, 2015) with the R

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

software. As for Accuracy, a first model was built, and variables of interest: Group, Number, and either 1) Plural Type (9 levels, see table 3 above) or 2) Number of differences between singular and plural forms were added one after the other, to obtain the model that best fitted the data, using model comparisons based on the AIC scores of the models. Contrast coding for the analyses of RT for categorical predictors is similar as for the analyses of Accuracy.

Statistical modeling. Gaze data: GAMM

Since we are interested in the processing of short words, we focused our analyses on the time frame comprising fixations from 200 ms before the onset of the singular-plural difference to 1500 ms after this point. Fixation towards the target is the dependent variable for gaze data analyses, ranging from 0 (i.e., no fixation towards the target in a 100 ms timebin) to 1 (i.e., fixation only to the target in a 100 ms timebin) evolving over time. To model this non-linear evolution of Fixation towards the target over time, we used Generalized Additive Mixed-Effect Models (GAMM), with the *bam()* function, working with the *mgcv* (S. N. Wood, 2011) and *itsadug* (van Rij et al., 2020) packages on R. As for Accuracy and RT, models were built step-by-step by adding factors to the simplest model. By-Participant and by-Item random smooths were added to the model. First, Group and Number (singular vs. plural) were added as variables of interest in the first analyses. Then, we conducted four separate types of analyses: 1) Plural Types (9 different levels, see Table 3); 2) Type of contrasting cue (binary variables: Umlaut Contrast, Suffix Contrast, Voicing Contrast, see Table 3); 3) analyses of the effect of the Number of Cues available to mark plural (from 1 to 3); and 4) effect of Age, Vocabulary Development and Phonological Working Memory in children only.

Results

To ensure readability of this section, detailed steps and outputs of each model (for Accuracy, RT and gaze data) built in the study are given in supplementary materials (https://osf.io/8t3pe/).

Picture selection

Singular or Plural bias?

Our choice of plural types included a “no change” Plural Type, in which plural and singular are identical. For this Plural Type, participants’ decision is neither accurate, nor inaccurate, since both forms are strictly identical[1]. For this reason, we exclude these items from our analyses of Accuracy, RT and Gaze Fixation. We did analyze them, to examine whether participants preferred singular over plural or not. As shown on Figure 2, both groups of participants chose the picture depicting the singular form of the words: adults (M = 64.76%, SD = 20.32) and children (M = 65.00%, SD = 19.53), with no significant between-group difference ($t(59.96)=-0.44$, p = .660). This gives an indication of a bias

[1]Acoustic measurements (word duration) do not point to an observable difference between singular and plural forms in our set of stimuli.

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

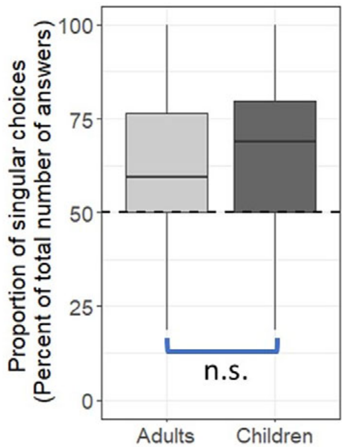

Figure 2. proportion of singular choices in both groups (% of the total number of answers) for the "no change" category.

towards the singular interpretation when no difference between singular and plural forms of the word is available.

Accuracy of selection

Overall accuracy of picture selection is very high in both adults (M = 99.45%, SD = 7.34%) and children (M = 95.02%, SD = 22.38%). Details of Accuracy per Group and Plural Type are given in Table 4. To analyze the factors influencing Accuracy, we built a GLMM with Accuracy as the binomial dependent variable and Participant and Item as random factors. The best-fitted model included Group (Adults: 1 vs. Children: -1) as the only fixed-effect factor. Number (Plural: 1 vs. Singular: -1) and Plural Type (nine levels) were not improving the model and were not included. Number of differences between singular and plural was improving the fit of the model and was included. The results indicate that the difference in Accuracy between Children and Adults (Adults: 1 vs. Children: -1) is small but significant ($\beta = 0.89$, $z = -3.88$, $p < .001$). Accuracy is significantly lower when only one cue is provided ($\beta = -0.31$, $z = -2.37$, $p < .05$).

Finally, we built a last GLMM to study the effects of individual factors on the Accuracy of the children. Participant and Item were included as random factors, and only Age was included in the model. Accuracy is higher in older children ($\beta = 0.04$, $z = 3.90$, $p < .001$). PPVT and NWR Scores were not improving the fit of the model and were therefore not included. These three possible fixed-effect factors are continuous variables.

Reaction times

Mean Reaction Times per Group for each Plural Type are given in Table 4 above. RT values were log-transformed to reduce the effects of the outliers. In our first model, RT is the continuous dependent variable, and Participant and Items are the random-effect factors. Group (Adults: 1 vs Children: -1), Number (Plural: 1 vs. Singular: -1) and Plural

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Table 4. Accuracy and Reaction Times per plural type and group

	Accuracy (in %)		Reaction Times (in ms)	
Plural type	Adults	Children	Adults	Children
Vowel change: umlaut	99.41 (7.64)	92.54 (26.30)	490.41 (405.69)	864.37 (879.26)
-/s/ suffix	99.22 (8.81)	93.86 (24.03)	481.33 (280.51)	886.16 (872.06)
-/ə/ suffix	99.61 (6.24)	94.08 (23.63)	439.10 (259.89)	804 (987.48)
-/n/ suffix	99.02 (9.84)	94.08 (23.63)	458.29 (227.73)	837.20 (1051.64)
Umlaut + -/ə/ suffix	99.61 (9.84)	96.49 (18.42)	431.47 (379.80)	767.59 (747.16)
Umlaut + -/ɐ/ suffix	99.41 (6.24)	95.61 (20.50)	451.53 (332.06)	888.67 (803.09)
Voiced stem-final C + /ə/ suffix	99.22 (7.64)	94.96 (21.91)	417.56 (277.73)	836.45 (747.09)
Voiced stem-final C + /ɐ/ suffix	99.80 (8.81)	96.05 (19.49)	442.55 (522.23)	789.62 (773.20)
Umlaut + Voiced stem-final C + /ə/ or /ɐ/	99.80 (4.42)	94.74 (22.35)	424.05 (360.91)	797.09 (788.27)

Type (nine levels) are the independent factors in the model, alone and in interactions. Results indicate that (1) children have a significantly longer RT, β = -0.28, t = -5.885, p < .001, (2) plural is processed significantly faster than singular, β = -0.52, t = -4.72, p < .001, and (3) there is a significantly longer RT for two plural types than for the other types – namely, the -/s/ and the -/n/ suffixes (-/s/ vs. mean RT: β = 0.12, t = 3.88, p < .001; -/n/ vs mean RT: β = 0.09, t = 2.83, p < .05) and a significantly shorter RT for one Plural Type than for the others – namely, umlaut + -/ə/ suffix (umlaut+ -/ə/ vs mean RT: β = -0.12, t = -3.75, p < .001). Finally, a pairwise comparison (Tukey correction) of the Plural Type x Group interaction indicate that the differences between Plural Types are mostly found in adults (see supplementary material for the details: https://osf.io/8t3pe/).

Starting from the simpler model with Group and Number as fixed-effect factors, we built a second LMM to analyze the effects of the number of contrasting cues. In this model, Group (Adults: 1 vs. Children: -1), Number (Plural: 1 vs. Singular: -1), Number of different cues between singular and plural types (one cues: 1, two cues: 0 or three cues: -1) alone and in interaction with Group are the fixed-effect factors. Children have a significantly higher RT than adults (β = -0.28, t = -5.99, p < .001), and RT is significantly longer with singular than plural target items (β = -0.05, t = -4.52, p < .001). The number of contrasting cues between singular and plural forms has a significant effect, alone and in interaction with the group.

Pairwise comparisons (Tukey correction) indicate that the RT of adults is significantly longer when plural is marked by a single cue than when by two cues (one vs. two cues: β = 0.14, t = 4.96, p < .001) or three cues (one vs. three cues: β = 0.20, t = 4.59, p < .001) with a non-significant difference between two and three cues. No significant effect of the number of different cues used to mark plural is observed in children, which explains the interaction effect. These results are presented in Figure 3.

Finally, we built a last LMM with child data only, to study the effects of Age, Vocabulary Development and Phonological Working Memory on RT. The best-fitted model includes Participant and Items as the random-effect factors, and Age and PPVT

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

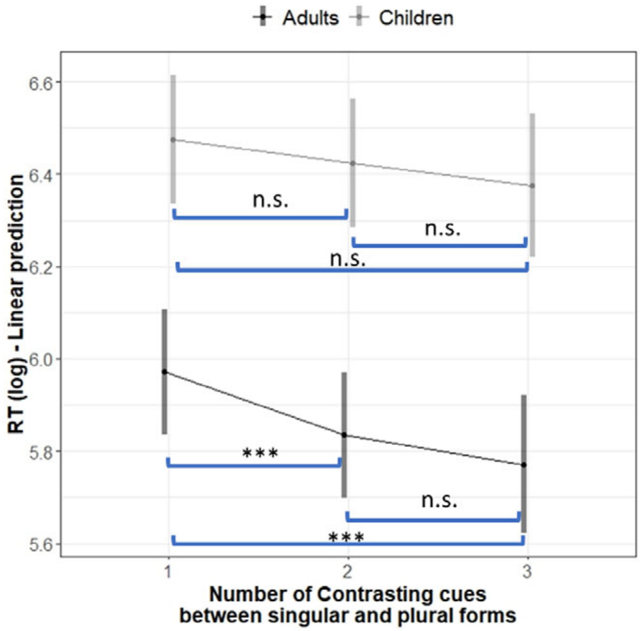

Figure 3. RT (log) for each group, and each number of different cues in pl-sg word pairs.

Score as the two continuous fixed-effect factors. RT is significantly lower in older children (β = -0.01, t = -4.90, p < .001), and in children with a higher PPVT Score (β = -0.02, t = -2.18, p < .01). Phonological Working Memory (i.e., NWR Scores) did not improve the fit of the model, and was not included.

Fixation times

Participants are presented two pictures and an audio token of the target word, which is either plural or singular. When the first cue to plural marking starts (0-point on our plots and analyses – corresponding to either the onset of a stem vowel, the onset of a contrasting voiced or voiceless stem-final consonant or the onset of a suffix, see Fig. 4-8 below and supplementary materials: https://osf.io/8t3pe/), participants start to direct their gaze towards one of the two pictures presented to them. The time between the start of the first cue and the end of the word is very short, (between 118 ms and 632 ms for plural items). The gaze fixation is thus very quickly oriented towards one picture, then reaches a plateau until the end of gaze recording (which corresponds to the time when the participant clicks on the chosen picture).

Group (Adults vs. children) and Number (Plural vs. Singular)

We built a first GAMM to examine whether children and adults have a similar fixation pattern and if the number of the word presented is influencing gaze fixation. The best-

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Table 5. Summary of the GAMM model used to study the effects of group and number

Parametric coefficients:	Estimate	Std error	*t*-value	Pr (>|t|)
(Intercept)	62.126	1.407	44.16	< .001 ***
Smooth terms:	**Edf**	**Ref.df**	***F***	***p*-value**
s(Time)	7.921	8.431	53.057	< .001 ***
s(Time):Group	8.231	8.905	11.416	< .001 ***
s(Time):Number	5.658	6.767	29.525	< .001 ***
s(Time, Participant)	259.242	557.000	1.959	< .001 ***
s(Time, Item):Group	498.094	1296.000	0.956	< .001 ***

"Group" is a variable modeling the difference between adults (1) and children (0) groups and "Number" is a variable modeling the difference between plural (1) and singular (0) target words.

fitted model includes Participant and Item as random-effect factors, and difference smooths for Group (adults vs. children) and Number (plural vs. singular target), modeling the non-linear difference of both factors. Table 5 presents the output of this model.

As shown in Table 5, there is a non-linear evolution of the difference in fixation towards the target by children and adults ($p < .001$). Figure 4 shows the evolution of the proportion of this fixation towards target per 100 ms timebins by both groups (fig. 4.A.) and the difference between these two groups. Adults direct their gaze earlier and reach a higher proportion of fixation towards the target picture than children very early on, around 75 ms after the time when the first cue to singular or plural for the word is to be heard.

Similarly, our statistical model indicates a non-linear evolution of the difference between plural and singular targets ($p < .001$): as shown in Figure 5, participants' gaze is directed towards plural target pictures earlier and with more amplitude than towards singular target pictures. This indicates a possible preference for plural target pictures.

Our model-selection procedure did not lead us to include an interaction of Group x Number, which indicates that the plural preference effect is indeed similar in both groups.

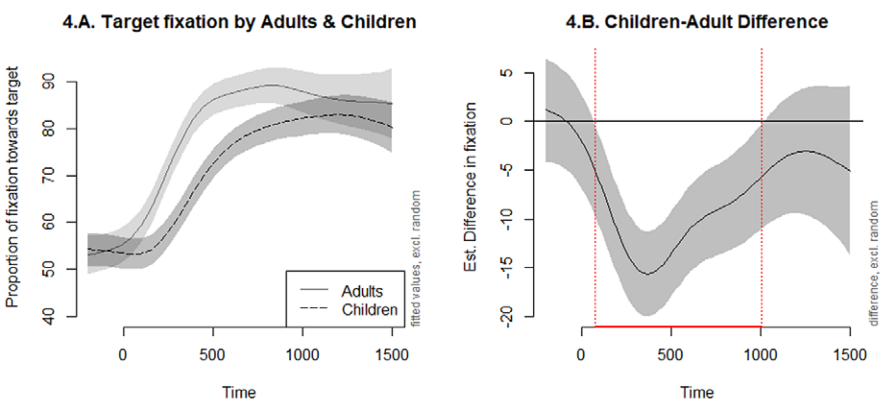

Figure 4. fixation towards target for adults and children (4.A.) and difference between children' and adults' fixation towards target, with children as the reference (4.B.).

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

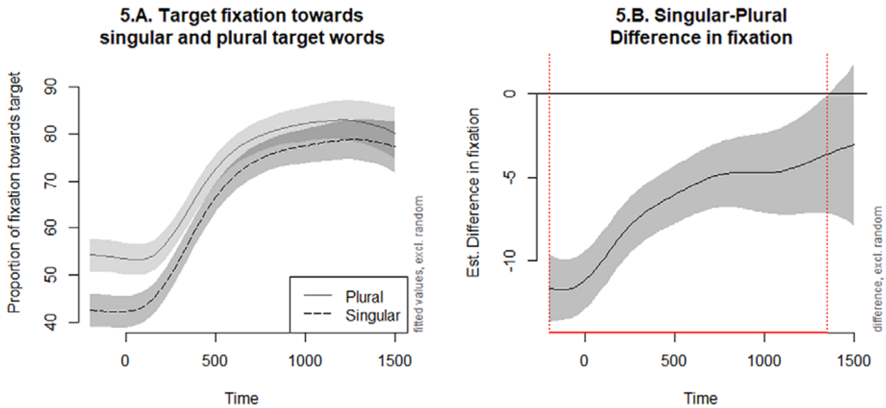

Figure 5. fixation towards plural and singular targets (5.A.) and difference between in fixation between singular and plural targets (5.B.).

In a further step, we built a new model by adding the mean Accuracy score of each participant across the entire task as a new dependent variable. Since Accuracy was lower in children compared to adults – but also lower in younger children (see Section "Accuracy of selection" above) – we are now considering all participants on a developmental continuum rather than from a categorical adult vs. child distinction. Our best-fitted model for this analysis included Participant and Item as random-effect factors, Group (adult vs. children), Number (Plural vs singular target) and Accuracy as a continuous factor. The results indicate that next to the already present effects of Group and Number, Accuracy is significantly affecting the evolution of Gaze Fixation over time ($p < .001$): the fixation is directed to the target more steeply and with more amplitude in participants with higher accuracy.

Plural Types (nine categories)

Starting from the model (including only Group and Number) of the previous section (Table 4), we added Plural Types (See Table 3 above) to create a new model. We then added separate composite smooths to model the Group x Plural Types and the Number x Plural Types interaction. The best-fitting model included Participant and Item as random-effect factors, and smooths for each Group x Plural Types and each Number x Plural Types. This allows us to explore the significance of between-Group and between-Number differences for each Plural Type. The main effects of Group and Number correspond to overall differences (e.g., Adults have an earlier and higher gaze fixation to the target than the children), and each interaction smooth gives more information as to when these differences are occurring exactly between Groups or between singular and plural targets. Visual exploration of all individual smooths and difference plots (See supplementary materials: https://osf.io/8t3pe/) are summarized in Table 6: for each Plural Type, we provide the location of between-Group and between-Number differences.

These results indicate that the difference in fixation for adults vs. children occurs very early with most Plural Types (earlier than 100 ms after the onset of the first cue to plural marking) and a little later for four Plural Types: these plural markings are those involving a vowel change, alone, or in combination with suffixes and/or subsequent voiced

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Table 6. Between-group and between-number differences for each Plural Types (latency in ms from the first singular-plural difference)

Plural Type	Differences in fixation between adults and children	Differences in fixation between singular and plural targets
Vowel change: umlaut	Earlier fixation for adults (from 298 ms)	Earlier fixation to plural targets (from 0)
-/s/ suffix	Earlier fixation for adults (from 195 ms)	n.s.
-/ə/ suffix	Earlier fixation for adults (from 75 ms)	n.s.
-/n/ suffix	Earlier fixation for adults (from 75 ms)	Earlier fixation to singular targets from 280 ms
Umlaut + -/ə/ suffix	Earlier fixation for adults (from 126 ms)	Earlier fixation to plural targets (from 0)
Umlaut + -/ɐ/ suffix	Earlier fixation for adults (from 212 ms)	Earlier fixation to plural targets (from 0)
Voiced stem-final C + /ə/ suffix	Earlier fixation for adults (from 75 ms)	Earlier fixation to plural targets (from 0)
Voiced stem-final C + /ɐ/ suffix	Earlier fixation for adults (from 92 ms)	Earlier fixation to plural targets (from 0)
Umlaut + Voiced stem-final C + /ə/ or /ɐ/	Earlier fixation for adults (from 126 ms)	Earlier fixation to plural targets (from 0)

realization of the stem-final consonant, and the -/s/ suffix alone. Note that, in these cases, the vowel change (Umlaut) always was the first cue to mark the plural and -/s/ is not found in combination with vowel change. Our results also indicate that plural targets are fixated earlier than singular ones for most Plural Types, except when only a suffix is provided: we find no significant difference or a preferred fixation towards the singular target occurring quite late (almost 300 ms) after the first singular-plural difference occurs.

Type of Cues available to mark plural (stem-medial vowel alternation, suffix and stem-final consonant voicing contrast)

One of our goals was to disentangle the contribution of phonetic, phonological and morphological information to the processing of plural marking: suffixation is a morphological process, stem-medial vowel change (i.e., umlaut) is a phonological process, and the subsequent voicing alternation is a phonetic adaptation (see our description of noun plural marking in German in the background section above). To this end, we reanalyzed Plural Types presented in the previous paragraph using vowel change (i.e., umlaut), voicing contrast and suffix as binary variables as presented in Table 3 above: plural marking relies on vowel change for four Plural Types, on suffixation for eight Plural Types and on voicing for three Plural Types (several Plural Types relying on more than one type of cues). For example, plural marked with "-/n/ suffix" is described as [-umlaut], [-voicing] and [+suffix]. We started again from the model presented in the previous section (see Table 5) and added the new possible binary fixed-effect factors one by one. Main effects of Group and Number are still present but will not be discussed further here.

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

The best-fitted model for these analyses includes Participant and Item as random-effect factors, and difference smooths for Group (adults vs. children), Number (plural vs. singular target), Umlaut (contrast vs. no contrast), Suffix (contrast vs. no contrast) and Voicing (contrast vs. no contrast) modeling the non-linear difference of these factors. The interactions of these three factors with Group and Number were also considered as fixed effect-factors. However, adding them did not improve the fit of the model and they were therefore not included. Figure 6 presents the evolution of the fixation toward the target when plural is marked with umlaut (e.g., Wald-Wälder, *forest*) or without umlaut contrast (e.g., Auto-Autos, *car)* and Figure 7 presents this evolution when plural is marked with a suffix (e.g., Wald-Wälder, *forest*) or without a suffix (e.g., Vogel-Vögel, *bird*).

For the Umlaut contrast, a visual inspection of the smooths (Figure 6.A.) and of the corresponding curve difference (Figure 6.B.) shows that there is a significant difference

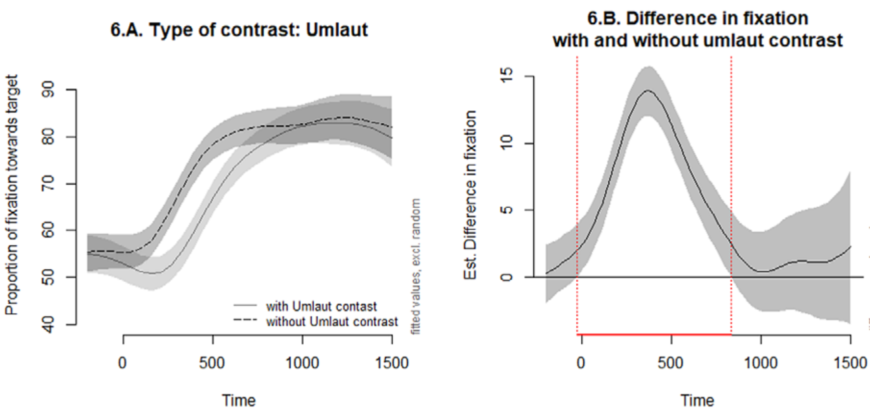

Figure 6. fixation towards the target when plural is marked with/without an umlaut contrast (6.B.) and difference between these two conditions (6.B.).

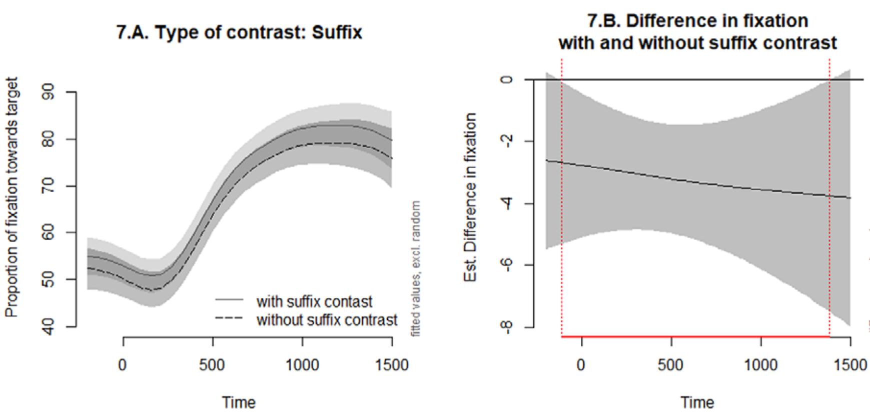

Figure 7. fixation towards the target when plural is marked with/without a suffix contrast (7.A.) and difference between these two conditions (7.B.).

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

between the two conditions: fixation occurs earlier when the plural is not marked with a vowel change, indicating a quicker access to the target when the word stem is identical in singular and plural forms of the word. Similarly, Figure 7 shows an earlier fixation when plural marking involves a suffix than when it does not. Similar analyses show no difference when plural is marked with a Voicing contrast. The differences in fixation coming from these three Types of Cues indicate that phonological (i.e., stem-medial vowel alternation) and morphological cues (i.e., suffixation) to plural marking affect processing of the markers, while this is not the case for the phonetic cue (i.e., voicing alternation on the stem-final consonant).

Number of Cues available to mark plural

Starting from the first GAMM including group and number again (see Table 3), we built a model to study the effect of the Number of Cues differing between singular and plural forms to mark plural and to study where between-Group and between-Number difference more precisely occur. This model included Participant and Item as random-effect factors. The interactions of the Number of cues with Group and Number did not improve the fit of the model and they were therefore not included. The model indicates an effect of the Number of Cues available to mark plural on the fixation towards the target. Representations of the evolution of fixation over time are presented in Figure 8.

Separate visual inspections of smooths differences show a significantly earlier fixation when one cue is available to mark plural than when two cues (Fig. 8.B.), or than when three cues are provided (Fig. 8.C.). These differences in fixation are marked quite early, around the onset of the first cue (1 vs. 2 cues provided), or shortly after this first cue occurs (1 vs. 3 cues provided). However, the difference when two or three cues are provided (Fig. 8.D.) occurs a little later, more than 200 ms after the onset of the first cue, and around the time of the onset of the second cue.

Children: effect of age, vocabulary development and phonological working memory

In order to understand how children specifically process plural marking, and what the effects of Age, Vocabulary Development and Phonological Working Memory are, we built a separate model with child data only. The best-fitted model includes Participant and Item, as random-effect factors, and the non-linear smooth difference for Number (plural vs. singular target). Our variable selection procedure led us to leave out Age, PPVT score and NWR score, indicating that they do not contribute to explaining the variation in gaze fixation. Note that this group of children was quite homogeneous with respect to PPVT and NWR scores, which might explain these results.

Discussion

The present study addressed the following question: which phonetic, phonological and morphological cues are used to process noun plural marking and how do children use this information in comparison to adults? To answer this, we used combined behavioral and eye-tracking paradigms, which allowed us to gain a fine-grained analysis of processing of plural marking.

Our results first indicated that when provided with no plural marking, i.e., the "no-change" plural type, both children and adults had a similar bias toward the singular

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

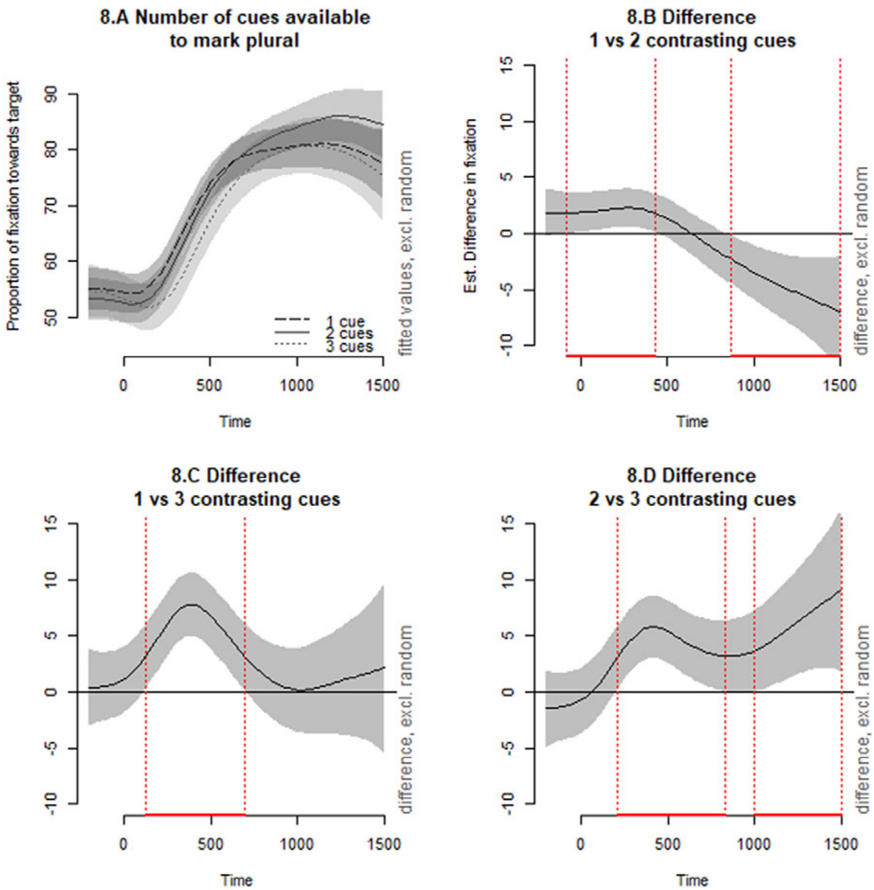

Figure 8. fixation towards target with the different number of cues available for plural marking (8.A.) and differences between number of cues (8.B., 8.C., 8.D.).

interpretation of the word. This is not unexpected as all items selected for the study were singular-dominant. For plural types involving formal differences between singular and plural, the results of RT (i.e., faster decision with a plural target) and gaze fixation (i.e., earlier and ampler fixation towards plural targets) showed overall a preference for the plural version of the item. This is also not surprising as more visual information was provided on the plural pictures than on the singular pictures (i.e., two objects instead of one): the eye drifts more easily towards the fuller surfaces and the decision is therefore quicker (see e.g., Carter & Luke, 2020 for a review of AOI-specific factors influencing gaze fixation).

Contribution of phonetic, phonological and morphological information to comprehension and processing of noun plural marking

The three different variables in our study (i.e., Accuracy, RT and fixation towards target) were affected differently by the types and combinations of cues to plural

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

marking. Accuracy was not found to be significantly affected by the different plural types, and RT was longer for two plural types (-/n/ and -/s/ suffixes) and shorter for only one other plural type (umlaut + -/ə/ suffix) than for the other types. RT gives a first indication that not all plural types are processed similarly. Our design allowed participants to respond only when the entire item had been heard, so that they would have to process all cues to plural marking before taking a decision. The change affecting the stem vowel was the first to occur, followed by the change affecting the stem-final consonant and finally suffixation. Therefore, we expected all plural types involving a stem vowel change to have shorter RTs and all plural types involving only a suffix to have longer RTs. However, this is only partially true, as RT is significantly shorter only for one plural type – namely, umlaut + -/ə/ suffix – than for the others including a stem vowel change. Furthermore, the longer RT with two plural types involving two different segmental suffixes – namely, -/s/ and -/n/ – but not with the syllabic suffix -/ə/ give us a first indication that processing is affected not only by the linguistic level (i.e., morphology, phonology or phonetics) but also to the form of the allomorphs (i.e., segmental vs. syllabic).

Fixation towards the target picture was earlier for plural types without an umlaut contrast than for those with an umlaut contrast, and for those with a suffix contrast than those without this contrast. A voicing contrast was not found to influence the gaze fixation towards target. This suggests that suffixation (morphological) is a more robust cue in plural processing than stem vowel or consonant changes (phonological and phonetic respectively). This is in line with the production study by Kauschke et al. (2011), which showed that children had more production errors for plural types involving stem vowel changes and less with those involving suffixation. This is also in line with production studies which showed that Hebrew-speaking children (Ravid & Schiff, 2009) and Arabic-speaking children (Saiegh-Haddad et al., 2012) had more difficulties with plural marking involving changing stems than non-changing stems.

Our main goal was to understand how different cues to plural marking were being processed by speakers (adults and children) – both alone and in combination with each other. We argued that not all plural types are processed similarly: phonetic, phonological and morphological cues are not used in processing plural marking to the same extent. Voicing does not seem to play a role in processing as gaze fixation did not differ depending on whether there is a voicing contrast or not, contrary to stem vowel change or suffixation. Phonetic information could thus be seen as less relevant for plural processing.

Accuracy was lower when plural was marked with only one cue than with two or three cues and RT was longer when one cue was used for plural marking than when two or three cues were available. Gaze fixation towards the target was earlier when only one cue was provided than when two and then three cues were provided, and the difference in gaze fixation towards the target between two and three cues was appearing later. This might indicate that when two or three cues are provided there is no effect of processing the first cue (i.e., a stem-final consonant or stem vowel change), but there is however an effect of processing the second cue (which is either a suffix if two cues are available or a stem-final consonant change if three cues are available). This could mean that the second cue is more relevant to identify plural forms of words than the first one, when more than one cue is provided. This different and earlier fixation when a suffix is involved as second and last cue (vs. voicing in case of three cues) further indicates that voicing alternation does not seem to be processed as a robust cue to plural marking.

Differences in noun plural marking between children and adults

Accuracy was very high for both groups, yet it was slightly lower for the children, who also had longer reaction times. For the adults, there was an effect of the number of cues to plural marking: with two and three cues available, the decision (i.e., RT) was quicker than with only one cue. This effect was not found for the children. Gaze fixation towards the target occurred earlier for the adults than for the children and with higher amplitude. This is true for all contrasts under consideration, with slight delays in the location of the between-group difference in gaze fixation when plural is marked with a -/s/ suffix or when a stem vowel change is involved. Furthermore, for the children, accuracy was higher in older children and RT was lower for older children and for children with a higher PPVT score. Looking further at the factors influencing gaze fixation, we found that neither age, nor lexical development, nor phonological working memory could explain variation in the children's gaze fixation pattern. However, gaze fixations of all participants (adults and children) were interacting with the accuracy of their responses: the most accurate individuals (older children and adults) are also the most decidedly looking at the target picture. Taken together, between-group differences for all measurements (Accuracy, RT, Gaze) are qualitatively in line with previous studies on English and German. Younger English-speaking children processed the three types of suffixes differently: while the segmental -/s/ suffix was understood as a plural marking, it was not the case for the segmental -/z/ and for the syllabic -/ɪz/ allomorphs, which were understood as plural marking only around the age of three years (Davies et al., 2017, 2020). German-speaking children aged from six to ten years did not have adult-like representation of noun plural marking yet (Korecky-Kröll et al., 2012), as in our study. The children in our study are aged 5;01 to 11;10, which corresponds to the later stages of inflection acquisition, when the so-called U-shape curve is stabilizing to adult-like characteristics in German-speaking children (Kauschke et al., 2011 for production; Korecky-Kröll et al., 2012 for comprehension). Our results thus show that children become more accurate at the end of the age range under consideration, and that a higher accuracy is a predictor of adult-like processing as shown in gaze fixation data.

Our results can be discussed regarding the following aspects of plural marking that – based on the literature – could play a role in children's processing of plural marking: regularity, predictability, productivity, and salience. Specifically, we hypothesized that salience of the individual morphemes and segmental information would influence processing of the plural types, where most salient markers are processed more similar in children and adults.

The most regular markers – namely, the -/s/ and -/n/ suffixes – are also considered to be default markers (Clahsen, 1999; Penke & Krause, 2002), and therefore were expected to show the smallest variation in processing between groups, that is, adult-like processing patterns in children. Our results however do not show a clear preference for these suffixes compared to other markers across measurements: no difference in accuracy is observed, RT is in fact longer for these two most regular markers, and between-group difference in gaze fixation is much earlier for the -/n/ suffix than for the -/s/ suffix. Regularity thus does not seem to be the strongest explanation in processing plural marking.

Predictability on the other hand seems to be playing a clearer role: the most predictable plural types are processed more quickly in our study, as indicated by an earlier fixation when a suffix is used to mark plural, compared to when an umlaut is used to mark plural. This fits the description of predictability given by Laaha et al. (2015), Albirini (2015), and Ravid and Schiff (2009): plurals with suffixes are more predictable than those with stem vowel changes.

The role of productivity can then be observed when comparing the different plural types: the between-group differences in gaze fixation starts later with the -/s/ suffix than for the -/ə/ and -/n/ suffixes. The most productive plural type in German is the -/s/ suffix and the least productive plural types include an umlaut (see Laaha et al., 2006, 2015). Our results are indicative of more adult-like comprehension patterns for the most productive types, and more different patterns between groups for the least productive types. Productivity can therefore be considered as a factor in the acquisition of plural marking.

Finally, we proposed salience of segments and morphemes as a predictor of plural comprehension and acquisition. Our definition of salience was based on both a phonetic (e.g., Clements, 2009) and a phonological approach (e.g., sonority, Selkirk, 1984). We expected vowel changes to be more salient than vocalic suffixes -/ə/ and -/ɐ/, and than consonantal suffixes -/s/ and -/n/. We also expected a different processing of voiced and voiceless consonants and thus a contribution of the voicing contrast to the morphemes' salience. This is not entirely what we found: our definition of the morphemes' salience was based on their segments' phonetic and phonological properties, which do not explain all variation in processing. Instead, the location of the change (word-internal-vowel changes or word-final-suffixes) seems to contribute to the morphemes' salience more than their sole phonetic or phonological properties. Furthermore, the difference in processing syllabic and segmental suffixes in our study is the opposite as in the study by Davies et al. (2020), indicating that the salience of plural morphemes is not equal cross-linguistically, and that in German, syllabic suffixes are more robust plural markings.

Limitations and future research

German provides an interesting system of plural marking, allowing for manipulating several types of cues alone or in combination. Where English plural marking only has three allomorphs, German plural marking is much more complex and relies on different types of cues. However, our design did not allow for paired comparisons of individual suffixes, as we used different combinations of suffixes with umlaut or voicing contrasts. Starting from this study, further work could for instance more specifically examine the effect of syllabicity on processing inflection. This could be done by studying processing differences between syllabic -/ən/ and segmental -/n/ suffixes, or between two different syllabic -/ə/ and -/ɐ/ suffixes, or between two different segmental -/n/ and -/s/ suffixes. Furthermore, our results indicated that even at a relatively late stage of language acquisition, children still do not show adult-like processing abilities with respect to plural marking.

Conclusion

This study contributes to the understanding that phonetic cues to inflection (i.e., plural marking) are less relevant for its processing than morpho-phonological ones and that having more cues available does not necessarily mean easier processing of inflection. It further shows that children from ages six to 11 still have not reached adult-like processing at that age. Finally, it underlines the importance of exploring languages with complex systems that rely on a variety of mechanisms and auditory cues, to have a more detailed view of how inflection is processed in acquisition.

https://doi.org/10.1017/S0305000923000521 Published online by Cambridge University Press

Acknowledgments. This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 - Project ID 390895286.

Competing interest. The authors declare none.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.

Albirini, A. (2015). Factors affecting the acquisition of plural morphology in Jordanian Arabic. *Journal of child language*, **42**(4), 734–762. doi: 10.1017/S0305000914000270

Baayen, R. H., **Piepenbrock, R.**, & **van Rijn, H.** (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia: Linguistic Data Consortium.

Bates, D., **Maechler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1–48.

Behrens, H. (2002). Learning multiple regularities: Evidence from overgeneralization errors in the German plural. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (Vol **1**. pp. 72–83). Cascadilla Press.

Behrens, H. (2011). Cues to form and function in the acquisition of German number and case inflection. In E. V. Clark & I. Arnon (Eds.), *Experience, Variation, and Generalization: Learning a First Language* (pp. 35–51). Amsterdam, Philadelphia: Benjamins.

Berko, J. (1958). The child's learning of English morphology. *Word*, **14**(2-3), 150–177.

Boersma, P., & **Weenink, D.** (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.23, retrieved 28 September 2020 from http://www.praat.org/

Carter, T., & **Luke, S.** (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, **155**. 49–62.

Chiat, S. (2015). Non-word repetition. In **S. Armon-Lotem**, **J. de Jong**, & **N. Meir** (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*. (pp. 125–150). Bristol, United Kingdom: Multilingual Matters.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and brain sciences*, **22**(6), 991–1013. doi: 10.1017/S0140525X99002228

Clahsen, H., **Rothweiler, M.**, **Woest, A.**, & **Marcus, G. F.** (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, **45**(3), 225–255. doi: 10.1016/0010-0277(92)90018-D

Clements, G. N. (2009). Does sonority have a phonetic basis? In E. Raimy. & C. Cairns (Eds.), *Contemporary views on architecture and representations in phonology*, 165–175. MIT Press. doi: 10.7551/mitpress/9780262182706.003.0007

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics (6th ed.)*. Oxford: Blackwell. Doi: 10.1002/9781444302776

Davies, B., **Rattanasone, N. X.**, & **Demuth, K.** (2017). Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects. *Language Learning and Development*, **13**(1), 38–53. doi: 10.1080/15475441.2016.1219257

Davies, B., **Rattanasone, N. X.**, & **Demuth, K.** (2020). Acquiring the last plural: Morphophonological effects on the comprehension of/-əz. *Language Learning and Development*, **16**(2), 161–179. doi: 10.1080/15475441.2020.1717956

Davies, B., **Rattanasone, N. X.**, **Schembri, T.**, & **Demuth, K.** (2019). Preschoolers' developing comprehension of the plural: The effects of number and allomorphic variation. *Journal of Experimental Child Psychology*, **185**, 95–108. doi: 10.1016/j.jecp.2019.04.015

Hinnerichs, J. (2016). *Spezifische Sprachentwicklungsstörungen bei Kindern mit Deutsch als Zweitsprache Zur Eignung des Nachsprechens von quasi-universalen Nichtwörtern als Differentialdiagnostik.* Unpublished Masters thesis.

Institut für Deutsche Sprache. (2009). *Korpusbasierte Grundformenliste DeReWo, v 40000g-2009-12-31-0.1, mit Benutzerdokumentation* [Corpus-based lexeme list DeReWo, v-40000g-2009-12-31-0.1, with user documentation].

Kauschke, C., **Kurth, A.**, & **Domahs, U.** (2011). Acquisition of German noun plurals in typically developing children and children with specific language impairment. *Child Development Research*, **2011**, article 718925. doi: 10.1155/2011/718925

Köpcke, K. M. (1988). Schemas in German plural formation. *Lingua*, **74**(4), 303–335. doi: 10.1016/0024-3841(88)90064-2

Korecky-Kröll, K., **Libben, G.**, **Stempfer, N.**, **Wiesinger, J.**, **Reinisch, E.**, **Bertl, J.**, & **Dressler, W. U.** (2012). Helping a crocodile to learn German plurals: Children's online judgment of actual, potential and illegal plural forms. *Morphology*, **22**(1), 35–65. doi: 10.1007/s11525-011-9191-8

Krech, E. M., **Stock, E.**, **Hirschfeld, U.**, & **Anders, L. C.** (2009). *Deutsches Ausspracheworterbuch*. De Gruyter Mouton.

Laaha, S., **Ravid, D.**, **Korecky-Kröll, K.**, **Laaha, G.**, & **Dressler, W. U.** (2006). Early noun plurals in German: regularity, productivity or default?. *Journal of Child Language*, **33**(2), 271–302. doi: 10.1017/S0305000906007379

Laaha, S., **Blineder, M.**, & **Gillis, S.** (2015). Noun plural production in preschoolers with early cochlear implantation: an experimental study of Dutch and German. *International Journal of Pediatric Otorhinolaryngology*, **79**(4), 561–569. doi: 10.1016/j.ijporl.2015.01.029

Lenhard, A., **Lenhard, W.**, **Segerer, R.**, & **Suggate, S.** (2015). *Peabody picture vocabulary test-4. Ausgabe*. Pearson.

Marcus, G. F. (1995). Children's overregularization of English plurals: a quantitative analysis. *Journal of child language*, **22**(2), 447–459. doi: 10.1017/S0305000900009879

Penke, M., & **Krause, M.** (2002). German noun plurals: A challenge to the dual-mechanism model. *Brain and language*, **81**(1-3), 303–311.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/

Ravid, D., & **Schiff, R.** (2009). Morphophonological categories of noun plurals in Hebrew: A developmental study. *Linguistics* **47**(1), 45–63. doi: 10.1515/LING.2009.002

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/

Saiegh-Haddad, E., **Hadieh, A.**, & **Ravid, D.** (2012). Acquiring noun plurals in Palestinian Arabic: Morphology, familiarity, and pattern frequency. *Language learning*, **62**(4), 1079–1109. doi: 10.1111/j.1467-9922.2012.00727.x

Selkirk, E. (1984). On the Major Class Features and Syllable Theory, in M. Aronoff and R. T. Oehrle (eds.), *Language Sound Structure*, MIT Press, Cambridge, Mass., pp. 107–136.

van Rij, J., **Wieling, M.**, **Baayen, R.**, & **van Rijn, H.** (2020). Itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs (R package version 2.4).

Wiese, R. (2009). The grammar and typology of plural noun inflection in varieties of German. *The Journal of Comparative Germanic Linguistics*, **12**(2), 137–173. doi: 10.1007/s10828-009-9030-z

Wilson, J., **Hartley, A.**, **Sharoff, S.**, & **Stephenson, P.** (2010, November). Advanced Corpus Solutions for Humanities Researchers. In *Proceedings of PACLIC 24*, Sendai, Japan. pp. 769–778.

Wood, J. N., **Kouider, S.**, & **Carey, S.** (2009). Acquisition of singular-plural morphology. *Developmental psychology*, **45**(1), 202. doi: 10.1037/a0014432

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, **73**(1), 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

Zaretsky, E., **Lange, B. P.**, **Euler, H. A.**, & **Neumann, K.** (2013). Acquisition of German pluralization rules in monolingual and multilingual children. *Studies in Second Language Learning and Teaching*, **3**(4), 551–580. doi: 10.14746/ssllt.2013.3.4.6

Zaretsky, E., **Lange, B. P.**, **Euler, H. A.**, & **Neumann, K.** (2016). Factors considered and ignored in plural acquisition: Frequency rules?. *Language and Cognition*, **8**(2), 283–313. doi: 10.1017/langcog.2014.51

Appendices

Appendix 1: Items' Lemma frequency for each plural type

Plural type	Mean lemma frequency (SD) in IPM (Instances per Million Words)
No change	22.45 (31.59)
Vowel change: umlaut	12.46 (10.42)
-/s/ suffix	33.40 (41.75)
-/ə/ suffix	27.49 (16.67)
-/n/ suffix	10.69 (7.56)
Umlaut + -/ə/ suffix	32.00 (21.06)
Umlaut + -/ɐ/ suffix	72.06 (135.87)
Voiced stem-final C + /ə/ suffix	38.78 (35.82)
Voiced stem-final C + /ɐ/ suffix	267.79 (331.90)
Umlaut + Voiced stem-final C + /ə/ or /ɐ/	29.09 (28.95)

Cite this article: Grandon, B., Schlechtweg, M., & **Ruigendijk, E.** (2024). Processing of noun plural marking in German-speaking children: an eye-tracking study. *Journal of Child Language* **51,** 1450–1477, https://doi.org/10.1017/S0305000923000521