German nominal number interpretation in an impaired mental lexicon: A naive discriminative learning perspective

(final version for *The Mental Lexicon*)

Ingo Plag, Maria Heitmeier & Frank Domahs

February 29, 2024

Abstract

There is an ongoing debate on how speakers and listeners process and interpret information in a morphological system that is very complex and not very transparent. A well-known test case is the German nominal number system. In this paper we employ discriminative learning (e.g. Ramscar & Yarlett 2007; Baayen et al. 2011, 2019) to test whether discriminative learning networks can be used to better understand the processing of German number. We analysed behavioral data obtained from a patient with primary progressive aphasia (Domahs et al., 2017), and the unimpaired system. We tested a model that implements the traditional cues borrowed from the schema approach (Köpcke 1988, 1993; Köpcke et al. 2021), and compare it to a model that uses segmental and phonotactic information only. Our results for the unimpaired system demonstrate that a model based on only biphones as cues is better able to predict the number of a given word-form than a model using structural phonological cues. We also tested whether a discriminative learning model can predict the number decisions by the aphasic patient. The results demonstrate that a biphone-based discriminative model trained on the patient's responses is superior to a structure-based model in approximating the patient's behavior.

1 Introduction

There is an ongoing debate on how speakers and listeners process and interpret form information in a morphological system that is very complex and not very transparent. A well-known test case in the debates about the nature of morphological and lexical organization is the German nominal number system. Beyond the inflection on the determiner or the adjective, number may be marked on the noun's stem vowel (by umlaut) and/or at its end (suffixes *-[e]n, -er, -e, -s, zero*). Other cues to number have been argued to include gender and the prosodic pattern of the noun (Wiese 2000, 2009), while semantics (e.g., animacy) is typically taken to play a rather minor role (see Eisenberg & Fuhrhop 2020, 171–176 for an overview).

There are two kinds of approaches that have transcended the words-and-rules approach of earlier times by developing more gradient, probabilistic views of the German nominal number system. The first is the schema approach, which was mainly developed by Köpcke and colleagues (Köpcke 1988, 1993; Köpcke et al. 2021). The second kind of approaches is computational and has seen the implementation of many different algorithms and architectures, from analogy to deep learning (e.g. Anderson & Lebiere 1998; Hahn & Nakisa 2000; Daelemans 2002; Wulf 2002; Daelemans et al. 2007; Rosen 2022; Buch 2011; McCurdy et al. 2020a; Beser 2021; Dankers et al. 2021).

In this paper we want to present and explore a third type of approach, discriminative learning (e.g. Rescorla & Wagner 1972; Rescorla 1988a,b; Ramscar & Yarlett 2007; Baayen et al. 2011). This approach can also be computationally implemented, and measures from discriminative learning networks have been successfully employed in morphological research to predict inflectional and derivational forms (Tomaschek & Ramscar 2022; van de Vijver & Uwambayinema 2022; Nieder et al. 2022), stress assignment to simplex and complex words (Arndt-Lappe et al. 2022; Tomaschek et al. 2023), the acoustic duration of suffixes (Tomaschek et al. 2021) and participants' reaction times and decisions in experiments investigating morphological processing (Baayen et al. 2011; Chuang et al. 2020; Heitmeier et al. 2023, 2024).

In the present study we have implemented discriminative learning networks for two purposes. First, we wanted to investigate the role of phonological cues in the German number system, testing a model that implements more traditional structural cues borrowed from the schema approach, and compare it to a model that uses sequences of segments only. This addresses the role of phonological information in inflectional systems.

Second, we wanted to see whether it is possible to use a discriminative learning model to predict the number decisions made by an aphasic patient when confronted with a given singular or plural form. In addition, we explored whether the discriminative learning network can be used to better understand the nature of the patient's problems with her processing of plural and singular nouns. Patterns of performance in patients with brain damage have been frequently investigated as a window into linguistic representations and processing. In the case of German grammatical number, studies using production tasks found evidence for a role of frequency - at least in irregular inflection, which is predominant in German number (Penke & Krause 2002), an influence of singular/plural dominance (Biedermann et al. 2018) as well as a general singular advantage (Biedermann et al. 2018), while the impact of morphological regularity is still a subject of debate. In a study investigating the comprehension of morphological inflection, Rath et al. (2015) found better performance for grammatical number compared to gender for most patients, which they argued to reflect a stronger supporting influence of semantic representations in the former compared to the latter.

The data for the present study were taken from Domahs et al. (2017). Their comprehension study aimed at demonstrating the influence of phonological cues on the interpretation of German noun phrases for grammatical number in a patient with an impaired mental lexicon. The authors found that the accuracy of the patient's decisions was influenced by the degree to which stimuli conformed to the prototypical, phonologically defined, schema for plural nouns (Köpcke 1988, 1993; Köpcke et al. 2021). However, the role of individual phonological segments or biphones – although in principle compatible with the schema account – has not been assessed.

The presentation of our study is structured as follows. In the next section we will discuss previous approaches to German nominal number, introduce discriminative learning networks and discuss how they can be implemented to investigate the nature of a morphological system. Section 3 explains our methodology, Section 4 presents the results. Section 5 discusses our findings and the implications for morphological theory.

2 Approaches to German nominal number

We will first present the schema account, then discuss computational models from a general perspective before zooming in on the specific model we implement in this paper, discriminative learning. Words-and-rules models (e.g. Marcus et al. 1995; Hahn & Nakisa 2000) have had a strong influence for a while but can no longer be considered state-of-the-art (compare the discussion in, for example, Pirrelli et al. 2021; Arndt-Lappe & Ernestus 2021; Fábregas & Penke 2021), which is the reason why we do not discuss them any further.

2.1 Schemas

Schemas are generalized abstractions over word forms that are mapped to a certain grammatical function, e.g., nominative plural (Köpcke & Wecker 2017). These abstractions can be seen as constellations of individual cues that together signal singular or plural in a probabilistic fashion. There are two kinds of schemas, called 'first-order schemas' and 'second-order schemas' (Köpcke et al. 2021). First-order schemas describe the properties of word forms in an output-oriented fashion, i.e. without relating to other forms within the same paradigm. Examples of first-order schemas for plural nouns would be 'Xe', as instantiated by *Hunde* 'dogs', or 'X-Umlaut-e' (e.g. *Bäume* 'trees'). Second-order schemas, in contrast, encapsulate paradigmatic relations (Hilpert 2019; Köpcke 1988). In the case of grammatical number, second-order schemas relate two first-order schemas to each other, which are mapped to the functions singular and plural, respectively. For instance, the pairing of the first-order schema for singular nouns 'die Xe' (as in *die Katze* 'the cat') with the first-order schema for plurals 'die Xen' (as in *die Katzen* 'the cats') forms a second-order schema.

Note that the formal cues that have been considered in the schema account of the German number system show different degrees of abstraction. They range from purely structural cues like the number of syllables (abstracting from their specific segmental makeup) over classes of phonemes (e.g. umlaut) to individual phonemes that (may) represent suffixes (e.g. *-e* or *-s*). The phonemic makeup of a whole word, as represented, for instance, by sequences of biphones, has not been discussed so far. The consideration of biphone sequences as formal cues as such does not seem incompatible with a schema account, though.

In the schema approach, the strength of an individual cue to signal singular or plural depends on factors related to its distribution in language use, as well as perceptual factors: High token frequency should lead to entrenchment, i.e., robust memory encoding of a specific form (Behrens 2009; Diessel 2019), while high type frequency should favor the generalization and abstraction of schemas on the basis of many individual forms (Köpcke et al. 2021; see below). In their specific ways, both high token and high type frequency facilitate learning and processing of form-meaning relationships.

The so-called 'validity' of a cue refers to the frequency in which it is used to express a certain grammatical function (e.g., plural) as opposed to other grammatical functions (e.g., singular). Given the ubiquity of syncretism, the same formal cue may be related to different grammatical functions. For example, word-final *-e* may occur in singular (e.g., *Büchse*) as well as plural nouns (e.g., *Füchse*). Moreover, final *-e* is also used for the marking of case across the paradigm. High cue validity favors learning and processing (Pescuma et al. 2021). One phenomenon that illustrates cue validity is zero marking, which often occurs with words whose singular form already ends with a string of sounds that is identical to that of available plural suffixes, e.g. *-en* or *-er* (e.g., $\textit{Rechen}_{\mathrm{Sg/Pl}}$, $\textit{Messer}_{\mathrm{Sg/Pl}}$). This decreases the cue validity of these suffixes as plural markers, as they also appear as endings in singular forms. Zero marking is also used for words ending in other schwa syllables like *-el* (e.g., $\textit{Pinsel}_{\mathrm{Sg/Pl}}$). Thus, given presence of this ending in plural forms, the ending *-el* can be a cue to plurality, even though it is not actually a suffix ('pseudo-suffix', Köpcke 1998).

The notion of 'salience' refers to the fact that not all cues are equally well perceivable (Köpcke 1998).[1] Segmentable cues and cues at the word's edges (e.g., suffixes) can be perceived more easily than non-segmentable or non-edge cues (e.g., umlaut) (for recent empirical evidence see Grandon et al. 2023). Complex cues, consisting of multiple segments, are more salient than cues consisting of only a single segment. Cues that contain a full vowel are more salient than those with a reduced vowel, which in turn are more salient than those with no vowel, followed by those with zero marking (Polišenská 2010). This is consistent with the assumption that cues forming a syllable on their own can be perceived more easily than non-syllabic cues – a factor which Köpcke (1998, 1993) called 'iconicity'. Higher salience of a cue has been shown to improve its learnability and processing in typical (Polišenská 2010; Simoens et al. 2017; Grandon et al. 2023) and atypical populations (Penke et al. 2016). Yet, the relative importance of these distributional and perceptual factors, their relative

[1] Note that this specific, perceptual sense of the term salience differs from other uses in linguistics (c.f. Boswijk & Coler 2020; Schmid & Günther 2016). In the schema approach (and also in some other approaches), salience is conceived as an inherently gradient variable that can be empirically determined (see MacLeod 2015).

impact on processing and their weights in interaction still remain unclear.

Studies using the schema approach have been quite successful in demonstrating that the formal cues to grammatical number in German operate probabilistically. Virtually no single cue is unambiguous or works in a deterministic way. However, so far the analyses of German number schemas are largely anecdotal or exemplary in nature, and a formal implementation is lacking.

2.2 Computational approaches

The second kind of approach to German nominal number involves the use of computational models. Studies in this domain apply various statistical and machine learning methods in order to correctly predict existing or novel wordforms, and to either explore potential human cognitive underpinnings or to only mimic human behavior.

Computational approaches have implemented a variety of architectures (see, for example, Anderson & Lebiere 1998; Hahn & Nakisa 2000; Daelemans 2002; Wulf 2002; Daelemans et al. 2007; Rosen 2022; Buch 2011; McCurdy et al. 2020a; Beser 2021; Dankers et al. 2021). While especially the more recent deep learning models perform quite successfully in general, they have certain drawbacks. First, with regard to the different phonological cues, the proposed models do not take all potential cues into account. Second, they are often not readily interpretable linguistically due to the nature of their architecture, for instance their including multiple hidden layers and their use of non-linear mapping functions. In this paper, we implement an alternative model, discriminative learning, with which we can address these concerns. The model is rather simple in its architecture and allows us to interpret the results, and how the system uses its information, from a linguistic perspective. Furthermore, the model is based on an established learning theory and therefore cognitively more plausible than deep neural networks.

2.3 Modeling nominal number with discriminative learning

Discriminative learning theory is a well-established theory of learning from cognitive psychology (e.g Rescorla 1988a; Pearce & Bouton 2001). The general cognitive mechanisms assumed in this theory have been shown to be able to model a number of important effects observed in both animal and human learning, for example the blocking effect (Kamin 1969) and the feature-label ordering effect (Ramscar et al. 2010).

Discriminative learning has recently been introduced to linguistics (see Plag 2018; Lieber 2021 for linguistic textbook introductions), and numerous studies have shown that discriminative learning can model many aspects of morphology. The central assumption of discriminative learning theory is that learning results from exposure to informative relations among events in the environment. These relations, or 'associations', can then be used to build representations of the world around us. The events associated with each other are called 'cues' and 'outcomes'. Associations between cues and outcomes (and the resulting representations) are constantly updated (i.e. they increase or decrease, representing learning and unlearning) based on new, informative experiences.

The association weights between cues and outcomes are computed mathematically using the so-called Rescorla-Wagner equations (Rescorla & Wagner 1972; Rescorla 1988a,b). The equations work in such a way that the association strength or 'weight' of an association increases with every time that a given cue (for instance, hearing the word *cat*) and a given outcome (for instance, seeing a cat) co-occur. Conversely, the weight decreases whenever a given cue occurs without that outcome.[2] Towards the end of the learning process a stable final state is asymptotically approached, with final association weights. The final association weights can be conceived as the activation of particular outcomes based on the training with all cues. These final association weights can also be directly computed by using the Danks equilibrium equations (Danks, 2003). The kind of model just described is known as 'Naive Discriminative Learning' (NDL, Baayen et al., 2011).

Figure 1 depicts an NDL network with the structural cues and singular/plural outcomes to be used in this study.

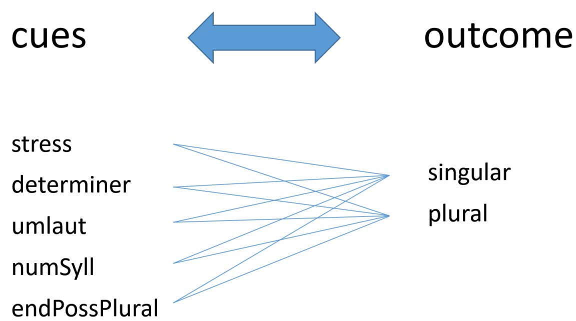

Figure 1: NDL network with the structural cues and the two number outcomes to be used in this study. numSyll = number of syllables, endPossPlural = word ends in possible plural suffix

3 Methods[3]

3.1 The patient

Domahs et al. (2017) reported on HT, a 56-year-old woman suffering from pri-

[2]In case an outcome occurs without any cue present, no learning can happen, no associations can be adjusted. For example, in an NDL model where cues are spoken words and outcomes are objects in a room described by the words, this would be a situation where there is a cat in the room, but no word is spoken.

[3]The data set and all scripts for the analyses are available at `https://osf.io/9578j`.

mary progressive aphasia. She was a native speaker of (Austrian) German. Further information on this patient can be found in previous publications with another focus of investigation (Domahs et al. 2006; Janssen & Domahs 2008). Based on a comprehensive battery of tests, the authors argued that HT's lexical knowledge was severely impaired. However, her application of general linguistic – including morphological – knowledge was largely preserved. Thus, HT had particular difficulty in all tasks in which word-specific lexical knowledge is required, such as naming, lexical decisions or assigning the correct (gender-marked) definite article to nouns, whereas linguistic tasks that do not require access to lexical knowledge (e.g., reading and writing pseudowords, nonverbal tasks to test semantic knowledge) were hardly affected at all (Domahs et al., 2006). For these reasons, this case was particularly interesting as it allowed researchers to investigate a human cognitive system without access to stored word-specific knowledge.

In the present study we focus on the results of a number decision task reported by Domahs et al. (2017), during which nominal phrases (consisting of a definite article and a simplex noun) were read to HT and she was asked to decide whether they referred to one or more entities. In the original study, HT's pattern of results was interpreted as evidence for a schema-based approach to the processing of grammatical number. On the one hand, her performance was in the significantly impaired range, but on the other hand, accuracy was clearly above chance level. Crucially, the correctness of her responses was systematically influenced by the cue strength of the respective schema: NPs forming good singular or plural schemas were better identified as singular or plural than NPs with weaker schemas. Each formal cue described as relevant in Köpcke's (1988, 1993) schema approach (article, syllable number, ending, and umlaut) as well as the noun's stress pattern had an influence on the certainty of HT's number decisions. Domahs et al. (2017) concluded that even in the case of largely lexicalized morphological paradigms such as the German number system, there may be formal cues that can be used in morphological processing.

3.2 Data set

In the present study, we analyze the data from the number decision task. We included all items for which CELEX provided a pronunciation and added that pronunciation. We eliminated the single word that had a pronunciation that involved a variable number of syllables (*Radio* 'radio', which can be pronounced with either two or three syllables). The resulting data set contained 924 nominal phrases, 463 of which were singular forms, 461 plural forms.[4] The patient took 464 plural and 460 singular decisions, with an 80 percent overlap (i.e. accuracy) between the target system and the patient's responses ($F1$=0.80, we computed

[4]The data contained eight lexemes which have identical singular and plural forms (*Bunker, Feuer, Finger, Keller, Kloster, Koffer, Teller, Ufer*). All of these end in segments that are identical one of the plural suffixes (i.e. *-er*). All of these lemmas are either masculine or neuter, so that, when presented with a preceding definite article, the stimulus is disambiguated as either plural or singular.

the $F1$ score to take both precision and recall into account).

Table 1 provides a description for each variable and its possible values. Table 2 gives the distribution of these variables.

Table 1: Cues and outcomes used in Domahs et al. (2017).

Variable	**Description**	**Values**
Outcomes		
NUMBER	number value of the stimulus word	`singular, plural`
RESPONSE	patient's response to the stimulus word	`singular, plural`
Cues		
STRESS	stress pattern of the word	`penult, other`
DETERMINER	determiner that accompanies the stimulus word	*die*, `other`
UMLAUT	presence of umlaut	`yes, no`
NUMSYLL	number of syllables	`1, 2`
ENDPOSSPLURAL	word ending is a possible plural suffix	`yes, no`

Table 2: Distribution of outcomes and cues. N=924

Outcomes		
NUMBER	`plural`	`singular`
	461	463
RESPONSE	`plural`	`singular`
	464	460
Cues		
ENDPOSSPLURAL	`no`	`yes`
	184	740
DETERMINER	*die*	`other`
	665	259
NUMSYLL	`1`	`2`
	199	725
UMLAUT	`no`	`yes`
	734	190
STRESS	`penult`	`other`
	681	243

3.3 Setting up the NDL models

The purpose of the NDL modeling was twofold. First we wanted to test how accurately the NDL-based association measures can predict whether a form is singular or plural. We call these models 'language models' as they model

German nouns as represented in healthy individuals. The outcome in these networks is the noun's NUMBER, i.e. plural or singular. Second, we wanted to test how accurately the NDL-based association measures can predict the patient's responses. We call these models 'patient models', as they model the impaired language system as represented in the patient. For these networks the outcome is the patient's RESPONSE.

For both the language models and the patient models we wanted to test two alternative cue structures. On the one hand we used the traditional structural phonological cues to implement so-called 'structure-based models'. On the other hand, we used biphones as cues, in the so-called 'biphone-based models'. Following other studies (e.g. Baayen et al. 2011; Tomaschek & Baayen 2017) we used biphones (and triphones, see below) as cues to see how mere segmental-phonotactic information, in contrast to more abstract structural-phonological information, is used by the model. The use of biphones is illustrated in Figure 2 with the word *Bach*, which is split up into the four biphones `#b`, `bA`, `Ax`, `x#` (we use the phonetic transcription symbols from CELEX, hash marks indicate word boundaries).

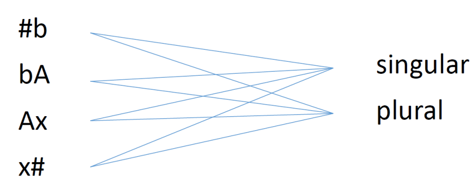

Figure 2: Partial NDL network with biphones as cues and singular and plural as outcomes. (Phonetic transcription symbols from CELEX, hash marks indicate word boundaries)

In addition to the structure-based models and the biphone-based models, we also implemented models in which both biphones and structural information were used ('all-cues models'). Table 3 lists the six types of models.

Table 3: NDL models for the language and the patient.

	name of model	**cues**	**outcome**
language models	structure-based	structural cues	number
	biphone-based	biphones	number
	all-cues	biphones and structural cues	number
patient models	structure-based	structural cues	response
	biphone-based	biphones	response
	all-cues	biphones and structural cues	response

As mentioned above, an NDL network computes the strength of association between cues and outcomes. As cues for the structure-based models we used the properties as coded by Domahs and colleagues, given in Table 1. These cues do not explicitly include grammatical gender. Although gender has been repeatedly demonstrated to be a strong cue to grammatical number in German in modeling studies (Daelemans (2002); McCurdy et al. (2020b); Dankers et al. (2021)), its use as cue seems inappropriate in the present study for three reasons. First, in German the gender of simplex nouns cannot be unambiguously identified in the input based on formal properties of the noun alone (Köpcke & Zubin (1996)). Rather, due to its largely non-transparent nature, gender is commonly assumed to be lexicalized and even known to be a notorious challenge in learning German (for a well known anecdotal report see Twain (1880)). Second, even if one assumes associations between a word's form or meaning and its gender at play, in a number decision task this gender cue would work much less directly compared to other cues to number (compare the processing path "form $\rightarrow$ (meaning) $\rightarrow$ gender $\rightarrow$ number" with the much more direct processing path for form based cues "form $\rightarrow$ number"). As a consequence, gender cannot serve as cue in number comprehension in the same way as formal properties do. Third, recall that the patient was presented with noun phrases including their definite article and that we do use the definite article as a cue in modelling. The distinction between *die* and *other* does not only contain number-related information, but provides partial information on gender, too, such that in nominative case the articles *der* and *das*, summarized as *other*, also signal masculine and neuter gender (in the singular), while *die* may be feminine singular as well as plural of all three gender types. Thus, gender information is not treated as a cue on its own, but partly contained in the definite article, which is used as a cue.

As outcomes we used NUMBER, with the values `singular, plural` to model the unimpaired lexicon, i.e. the language system. To model the patient's system we trained the model on the patient's responses.

Recall that associations between cues and outcomes are strengthened upon the presence of a particular cue and the presence of a particular outcome, while associations are weakened upon the presence of a particular cue and the absence of a particular outcome. We used the `ndl` package (Baayen et al., 2011) in R with the `estimateWeights`, `estimateActivations` and `NDLclassify` functions to compute the association weights between cues and outcomes using Danks' equilibrium equations (Danks, 2003). We also provided the functions with the words' frequencies, such that the equilibrium equations approximate the weights which would have been learned if each cue-outcome pair had been presented according to its frequency of occurrence. Frequencies were taken from CELEX (Baayen et al., 1996). For the cross-validations we used the package `Judiling` (Luo et al. 2023) with frequency-informed learning (Heitmeier et al., 2024) to match the implementation in `ndl` as much as possible[5].

The association weights allow us to implement two kinds of analyses. In one

[5] We had to use JudiLing for cross-validation since the `NDLClassify` function in `ndl` is not able to deal with missing form cues. Using it would therefore have led to a skew in the results where words with biphones which do not occur in any other word in the data could not have been used as held-out data.

kind of analysis we can look at how predictive the cues are for a given word. In this analysis, for each word one adds up the weights of the cue–outcome connections as instantiated in this word, thereby gauging the overall activation of singular and plural by that word-form. For instance, using all weights of all pertinent cues for the stimulus *die Achse* 'the axis', we obtain an activation weight of 0.91 for singular, and 0.09 for plural. To predict the number of a given form we now simply choose the higher of the two activation weights for the classification of the form as being either singular or plural. In the case of *die Achse* the predicted number is singular. In the second type of analysis we can investigate how predictive each (type of) cue is for singular and for plural. We will employ both analyses to model the language system and the patient's system.

4 Results

4.1 NDL: Overall performance

4.1.1 The language

We first look at the predictive power of the model for the classification of individual word-forms in the language. Using only the five structural properties STRESS, DETERMINER, UMLAUT, NUMSYLL and ENDPOSSPLURAL as cues, and NUMBER as outcome (the structure-based model), we achieve a moderate overall accuracy of 0.78 ($F1$=0.82). There is, however, a problem: Discriminating a plural form is easier than discriminating a singular form. While none of the plural forms is wrongly classified by the model as singular, 41 percent of the singulars are mis-classified as plurals. This means that the cues essentially encode properties of plural forms, and are not informative about singulars. Or, viewed from a different angle, roughly 40 percent of the singulars seem to have properties that are more associated with plurals than with singulars.[6]

Let us now turn to an alternative model, in which we do not use the structural information encoded in the above cues, but use only the segmental and phonotactic information encoded in biphones (the biphone-based model), or triphones. Fitting a network with only biphones as cues results in an accuracy of 0.93 ($F1$=0.94). 96 percent of the plurals are correctly predicted and 91 percent of the singulars. An alternative model using triphones reaches an accuracy of 0.96 ($F1$=0.96). However, the triphone model reaches a lower accuracy under leave-one-out cross-validation (0.74 compared to 0.78 for the biphone model; see below). Together with the fact that the biphone model has a much lower number of cues (1204 vs. 2888), the biphone model is therefore preferred.

[6] To assess the performance of the NDL classifier from a different perspective we also implemented a random forest analysis with the same structural predictors (using the `partykit` package and its `cforest()` function with default settings, Hothorn & Zeileis 2015). The $F1$ score of 0.83 was only slightly higher than that of the NDL model ($F1$=0.82), and the imbalance between the predictions for plural vs. singular nouns is also basically the same in the random forest model: 40 percent of the singulars are misclassified as plurals.

In order to check whether an NDL model using biphones and triphones is able to detect suffixes (as emergent properties), we took a closer look at the association weights of the individual biphones and triphones. We first computed the difference between the plural association and the singular association for each biphone. The mean difference is 0.20, with a range of [-3.5, 5.4] (for the patient model: [-3.52, 4.49]). The strongest association of a final sound with plural has final /a/, followed by final /s/ and final schwa. The associations of these individual final sounds with plural are, however, not very strong, with differences of 0.19, -0.07 and -0.16, respectively. Two of these, i.e. /s/ and schwa are indeed plural suffixes, but their associations might have been expected to be higher. One reason for that may be the simple fact that many singular forms end in segments that may represent plural suffixes. In other words, the final segment itself is not very discriminative.

What may also play a role in the association weights of the final sounds with plural is that it is not only the final segment, but a longer final string that is important for discrimination. This cannot be tested with biphones, however, because of way the coding of the end of a word is done. The word margin is treated as a signal, on a par with a phonological segment: the word-final biphone consists of the final sound and the margin (e.g. `x#` see figure 2). So, in the biphone model, we cannot check the role of the segment preceding the final sound. This can, however, be done with a triphone model. In the triphone model, associations of triphones with plural have a range of [-2.53, 2.0] with a mean of -0.10. Among the top ten triphones associated with plural, we find three triphones with the right margin of the word as last character. The one with the highest association, and indeed the second highest association of all 1444 triphones, is the one with schwa followed by /n/, followed by the word margin. The other two combinations of final segments among the top ten triphones are /as/ and /os/. These findings support the idea that an NDL model trained on sequences of biphones or triphones is able to detect the significance of strings that morphologists call ‘morphs’.

If we take both the structural features and the biphones as cues (the ‘all-cues’ model), the model predictions improve in accuracy from 0.93 ($F1$=0.94) of the biphone model to 0.98 ($F1$=0.98) of the all-cues model. This means that the structural information improves the model fit only slightly, and that the NDL model with the biphone information alone is already able to quite accurately discriminate singulars and plurals. Adding the higher-order structural information does not contribute much in addition.

The performance measures of the NDL models discussed so far are all based on testing the model on the data it is trained on. It is, however, necessary to investigate whether the model is able to deal with new, unseen forms. This is important especially with regard to the models using biphones, where the good performance could in principle be based solely on memorization. If the model still performs reasonably well in cross-validation, even if worse than when tested on the data it was trained on, this means that the model is able to generalize from existing knowledge to novel data.

To test this, we employed cross-validation using the leave-one-out procedure.

In this procedure, one item is withheld from the training set. The number value of this item is then predicted by the model.[7] Note that with the NDL model based on structural features we expect the cross-validation to show a very similar accuracy as the model that is tested on itself, as this model raises no suspicions concerning possible lexicalization effects, and because the structural predictors embody abstract properties that are easily applied to new data. The fit of the models (as gauged by acccuracy) are documented in Table 4.

Table 4: Accuracies of NDL models for the language.

Model	**Training**	**Cross-validation**
structure-based	0.78	0.79[8]
biphone-based	0.95	0.78
all-cues	0.99	0.89

We see that the accuracy of the models involving biphones suffer under cross-validation, but still reach satisfactory levels. This means that even the NDL model that is solely based on biphones is able to predict the number of unseen nouns using the associations of biphones with singular and plural. In contrast to the biphone-based model, the structure-based model does not suffer from cross-validation.

Next, we were interested in whether learning in the systems (based on structural cues, biphones, or both) depends particularly on high-frequency words, on low-frequency words or on all words equally. We therefore ran two simulations where we gradually left out more and more words in the training phase, and then tested how well the model was able to predict the number value for all words in the data set. In the first simulation, the parts of the data to be left out were randomly selected. In the second simulation, we eliminated words in ascending order of frequency, i.e. we first eliminated the words with the lowest

[7] We also experimented with 10-fold cross-validation where the data is split into ten equally sized parts, and tested on each of them after being trained on the remaining nine parts. However, we found two issues with this approach. First, leaving out 10% of the data makes it relatively likely that some of the biphone cues within in the 10% do not occur in the 90% training data. This means that the model is tested on many cues it has never encountered before, in contrast to both leave-one-out cross-validation and real life, where speakers encounter many more words than the approximately 924 nouns in this dataset. A second option would be to apply what Luo (2021) call 'careful split', where the held-out data is selected such that all biphone cues in the held-out data also occur in the training data. This method, however, prevents a full 10-fold cross validation, since not all items in the data can be tested. A second issue is that with 10-fold cross validation a reduction in accuracy may also be attributed to the smaller training data size. Therefore, it is unclear whether any reduction in accuracy is due to a lack of generalization ability of the model, or simply because it had less data to learn from. With leave-one-out cross validation both of these issues are alleviated.

[8] The somewhat unintuitive result that the cross-validation performance is higher than the training performance is due to the fact that training performance was computed with the Danks equations in `ndl` while cross-validations were computed with `JudiLing`. Though the two are approximating the same frequency-informed endstate of learning, slight differences due to their different mathematical derivations and computational implementations are to be expected.

frequency, and the last words to be eliminated were the words with the highest frequency. If there is a different effect of words with different frequencies, we should see a clear difference between the two simulations. If the learning of the systems does not depend on high- or low-frequency words in particular, both simulations should give similar results.

The simulations were done for the three types of model, i.e. the structure-based model, the biphone-based model and the all-cues model. The results of the simulations are shown in Figures 3 and 4. The size of the training set decreases from left to right, as shown on the x-axis. The y-axis gives the accuracy of the respective model. The graph can also be read from right to left, indicating the learning curve of the model, with the predictions being based on increasing training data.

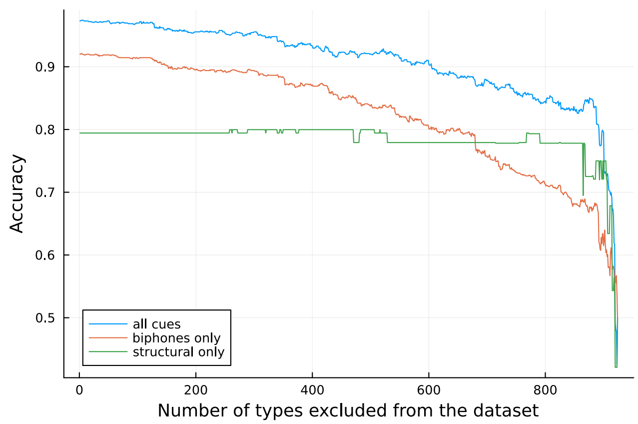

Figure 3: Cross-validation results for three different models under random selection of left-out data.

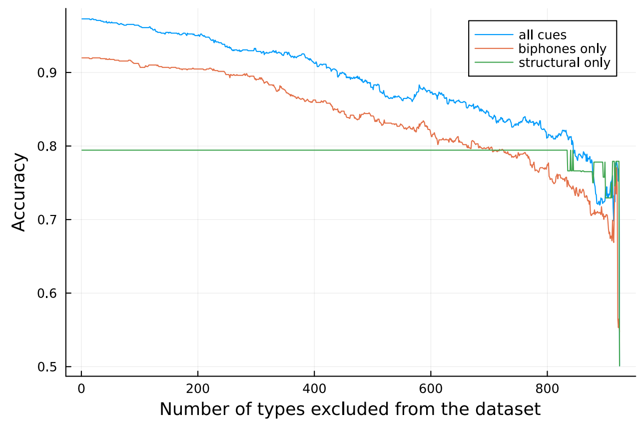

Figure 4: Cross-validation results for three different models under frequency-based selection of left-out data. Frequency increases from left to right.

Let us first look at what characterizes model behavior in both simulations. Quite expectedly, under both regimes, the performance of the biphone model and the all-cues model deteriorates with smaller training sets. With fewer training data, the model may not have seen all biphones, hence for these biphones the model has no associations to number yet, or the existing associations are still unstable due to the small number of learning events. The performance of the structure-based model is, however, very stable after a short initial learning curve over roughly the first 100 words (reading the graph from right to left). After that, no more learning happens and the predictions cannot improve. Given the worse overall performance and the undesired strong imbalance in the structure-based model between plural predictions and singular predictions, which is retained in all cross-validations, the structure-based model is much less appealing than the biphone-based model.

The differences between the two regimes are rather small, which means that token frequency does not seem to play a very strong role in establishing associations (though note that in both simulations, models were still trained in a frequency-informed way). In the frequency-based cross-validation, all curves are somewhat closer together at the right edge of the plot than in the random-order cross-validation.

4.1.2 The patient

We first trained a structure-based model on the five structural cues and the patient's responses as outcomes. Interestingly, this model reaches only an accuracy of 0.71 ($F1$=0.75), which means that, overall, the cues are less strongly associated with the patient's responses than with the nouns' actual number value. The predictions based on these associations are still much better than chance, however, which may suggest that, as concluded by the authors of the original study, the patient still relies to some extent on the phonological schemas for number.

But again, there is a striking imbalance between responses to plurals versus responses to singulars. Responses to plural forms are much more successfully predicted (0.88) than responses to singular forms (0.55) although there was no such imbalance in the responses of the patient. Overall, the structure-based NDL model does not seem to adequately reflect the patient's system.

We then trained a biphone-based model using only biphones as cues, and the patient's responses as outcomes. This model reached an accuracy of 95 percent ($F1$=0.95), with excellent balance between the prediction of plural and singular responses: 0.95 accuracy for plurals, 0.95 for singulars.

Let us compare the association weights of individual biphones with plural and singular in the language model with those of the patient model. Figure 5 shows the ranks of the association weights with plural for the final segments for the two models. Highest ranks are in the bottom left corner.

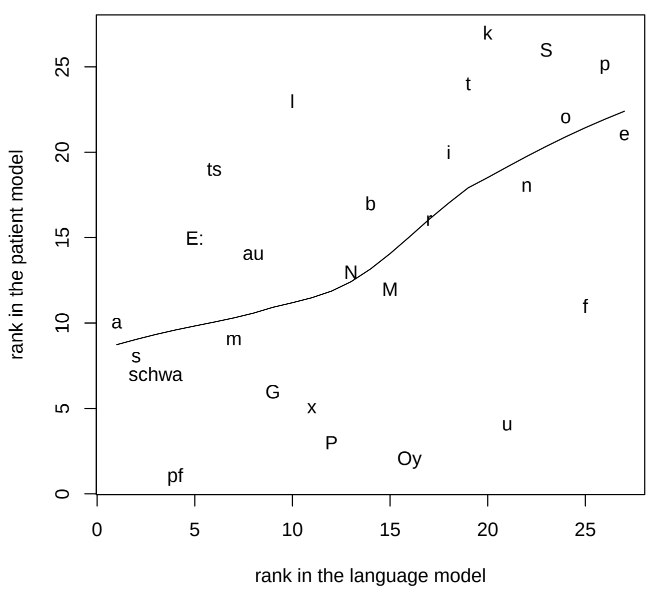

Figure 5: Correlation between the ranks of associations weights of final sounds with plural in the language model and in the patient model. (Some of the IPA symbols had to be recoded to allow processing by the statistics program.)

The line shows the trend in the data using a non-parametric scatterplot smoother. It can be seen that, quite expectedly, there are differences in the ranking of the association weights, but that, also quite expectedly, there is a clear and significant correlation of association weights across the two models (rho=0.51, p=0.007, Spearman). Notably, in both models, final segments that also feature as plural suffixes (or parts thereof) have high ranks (/s/ and schwa).

Furthermore, we explored how well the biphone-based NDL model trained on the patient's responses, i.e. the patient's system, is able to determine the correct number of the nouns. The patient's system reaches an overall accuracy of 78 percent in detecting the number of the given nouns. Plurals reached 77 percent, singulars 79 percent correct detection. These figures are very close to the patient's actual accuracies in the experiment (overall: 0.80, plurals: 0.80, singulars: 0.80).

To investigate the biphone model further, we compared whether the patient's actual correct and incorrect responses are the same as the patient's model's predicted responses. There is indeed a large overlap of 94 percent between the patient and the model (88 percent for incorrect responses, 95 percent for correct responses, 92 percent for plurals, 95 percent for singulars). This means that the biphone-based NDL model of the patient's system, in contrast to the structure-based model, is highly successful in predicting the patient's behavior.

A parallel analysis for the structure-based model shows that the structure-based model does not approximate the patient's system very well. There is only an overlap of 70 percent between the patient's actual responses and the predictions of the structure-based patient model (36 percent for incorrect responses, 78 percent for correct responses, 80 percent for plurals, 62 percent for singulars). The biphone-based NDL model thus is a much better approximation to the patient's system than the structure-based model.

We also fitted an all-cues model for the patient. The model reached the same level of accuracy as the biphone-only model (0.96), with balanced accuracy (0.96 plural, 0.96 singular).

Analogous to our analysis of the language system we also performed a leave-one-out cross-validation, with the results shown in Table 5.

Table 5: Accuracies of NDL models for the patient's responses.

Model	**training**	**cross-validation**
structure-based	0.71	0.71
biphone-based	0.95	0.71
all-cues	0.96	0.73

The performance of the NDL patient models with biphones show very good performance in predicting the number decision for words in the training set, but shows a larger decrease under cross-validation than we saw for the language models (patient: 0.95 vs. 0.71, language: 0.95 vs. 0.78). This means that the associations between the biphones (or the biphones plus the structural cues) and the two number values are well entrenched in the patient's system, but there are some problems generalizing to unseen forms.

4.2 NDL: The strength of the structural cues

The structural cues are not independent of each other. Figure 6 shows the correlations among the five cues. Apart from UMLAUT, all cues, quite expectedly, show rather strong inter-correlations. For instance, stress is highly dependent on the number of syllables, as only bisyllabic forms, but not monosyllables, may have penultimate stress.

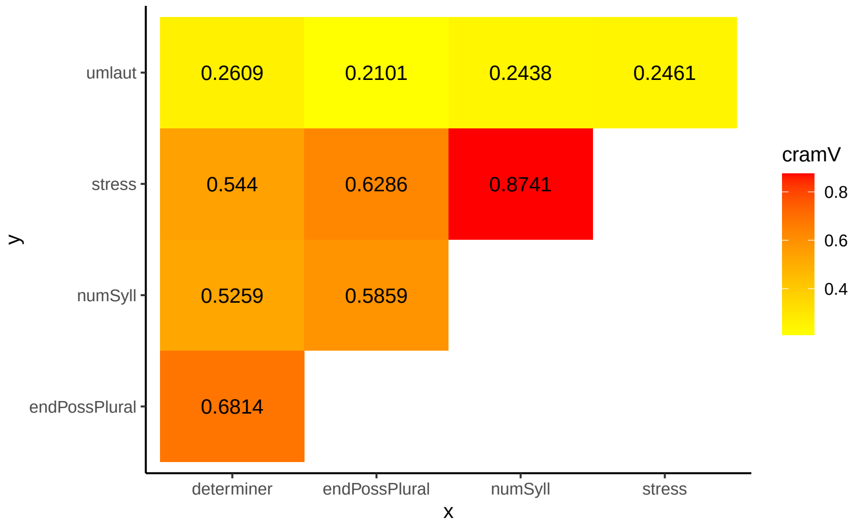

Figure 6: Correlations between structural cues. Numbers give Cramer's V (CramV); darker colors indicate stronger intercorrelations.

Given the non-independence of the cues, it is impossible to meaningfully interpret the work of the individual cues in a model that uses all cues. Like in multiple regression, the performance of such a model in predicting outcomes is interesting and valid, but the contributions of the individual cues are not very well interpretable due to collinearity (Friedman & Wall, 2005).

To assess the role of individual cues we adopted two strategies. First, we fitted five NDL models each with only one of the five structural cues and NUMBER as outcome. Second, we fitted models from which we took out one of the five cues and then gauged the decrease in performance quality vis-à-vis the model with all five cues by means of $F1$ scores. The activations for each individual cue for plural and singular are given in columns 3 and 4 of Table 6. Recall from above that the $F1$ score of the language model with all five cues was 0.82.

Table 6: Activations of plural and singular in the language models with only this cue (columns 3 and 4) and $F1$ scores of the full structure-based models without this cue. The rightmost column gives the decrease in the $F1$ score when compared to the $F1$ score of the language model with all five cues ($F1$=0.82).

cue	**value**	**plural**	**singular**	$F1$ **w/o this cue**	**decrease in** $F1$
STRESS	`penult`	0.60	0.40	0.82	0
	`other`	0.23	0.77		
DETERMINER	*die*	0.63	0.31	0.77	0.05
	`other`	0.00	1.00		
UMLAUT	`yes`	0.69	0.31	0.82	0
	`no`	0.45	0.55		
NUMSYLL	`1`	0.17	0.83	0.7	0.12
	`2`	0.59	0.41		
ENDPOSSPLURAL	`yes`	0.62	0.38	0.82	0
	`no`	0.00	1.00		

We can see in columns 3 and 4 that the activations of the individual cues play out quite as expected, reflecting the distribution of the cues in the data. Penultimate stress is more strongly associated with plural than with singular, other stress patterns more with singular. The determiner *die* is associated with plural, other determiners with singular. Umlaut is somewhat discriminative for plurals but not for singulars. Monosyllables are associated strongly with singular, while having two syllables is more associated with plural (even if weakly). Having an ending that could be a possible plural suffix is associated with plural, and having no such ending is strongly associated with singular. The $F1$ scores show that the performance of the model deteriorates when DETERMINER or NUMSYLL are not part of the model, while the loss of any one of the other three cues hardly affect the performance. This means that the kind of determiner and the number of syllables are the most important cues.

In Table 7 we list how much the structural cues activate plural and singular in the patient's models. The two rightmost columns give the differences to the language models.

Table 7: Activations of plural and singular by individual structural cues in the NDL patient models with single individual cues as only cue.

				Difference to language	
cue	**value**	**plural**	**singular**	**plural**	**singular**
STRESS	penult	0.62	0.38	0.02	-0.02
	other	0.18	0.82	-0.05	0.05
DETERMINER	*die*	0.63	0.37	0.00	0.06
	other	0.16	0.84	0.16	-0.16
UMLAUT	yes	0.46	0.55	-0.23	0.24
	no	0.68	0.32	0.23	-0.23
NUMSYLL	1	0.12	0.88	-0.05	0.05
	2	0.61	0.39	0.02	-0.02
ENDPOSSPLURAL	yes	0.59	0.41	-0.03	0.03
	no	0.14	0.86	0.14	-0.14

The two systems are quite similar in their activations, which is no surprise given the 80 percent overlap of the the patient's responses with the actual number of the words in the experiment. The most interesting and largest difference between patient and language is found with umlaut. The umlaut associations of the patient systems go into the opposite direction from that in the language.

5 Discussion

Let us summarize our main results. The structure-based NDL models both for the language and for the patient reached a moderate performance at best and were not able to detect singular forms reliably. In contrast, models based on only biphones as cues were very well able to predict the number of a given word-form and did not show any imbalance between predictions for singular and plural forms. Models using triphones or models combining structural with biphone cues achieved only very minor improvements. When applying cross-validation procedures, the biphone models still reached satisfactory levels of accuracy, while the models based on structural cues preserved their level of performance. A model trained on biphones as cues and the patient's responses as outcomes performed quite similar to the patient when deciding on the actual number of a given noun phrase - both in terms of her global error rate and concerning the response pattern for individual items. The definite article and the noun's number of syllables turned out as most relevant structural cues. Finally, associations of the structural cues with number values were quite similar for the language and the patient models except for the role of umlaut, where the patient model showed slightly stronger associations of umlaut with singular and of no umlaut with plural.

The generally moderate performance with structural cues was not expected from the point of view of the schema account. In particular, structural cues were poor in predicting the number value of singular forms. This weakens the appeal of this approach considerably.

A more detailed analysis of the role of the individual structural cues has shown that the determiner and the number of syllables are the strongest cues for grammatical number. While this result is compatible with more traditional accounts on grammatical number in German (Köpcke & Wecker, 2017; Köpcke et al., 2021; Wiese, 2000, 2009), the weaker role for (possible) plural suffixes appears surprising at first glance. However, as already demonstrated by Köpcke (1988, 1993), Köpcke et al. (2021), some forms have very poor validity, being used quite frequently both at the end of singular and plural nouns (e.g. *-e, -er*) and no single German suffix is unequivocal in this respect. A weak role of umlaut is expected in the schema approach given both the low validity and salience of this cue. This weak cue strength of umlaut in general may have contributed to the unexpected pattern of this cue in the patient data.

The exploration of models that do not use the structural cues from the schema account, but only biphones, has demonstrated that these models can successfully predict both singular and plural, even for novel forms. For held-out forms the performance of the biphone model was at the same level as the structure-based model, but yielded a balanced performance for singulars and plural forms. Given their more abstract nature of cues, the structural models did not suffer from cross-validation procedures in the same way as the biphone-models did. Overall, however, the biphone-based models were better able to approximate the language system and to predict both plurals and singulars quite reliably solely on the basis of phonotactic information.

With regard to the patient, the results play out even more favorably for the biphone-based models. The structure-based model was again empirically much less adequate, and yielded an imbalanced accuracy across the two number values. The biphone-based model, however, was able to simulate the patient's behavior quite successfully.

The biphone model has many more parameters than the structure-based model. It was shown, however, that over-fitting is not an issue. The cross-validations demonstrated that the biphone model is able to generalize to unseen data. The reader might wonder whether longer words profit from the fact that they have more biphones, hence supposedly more information concerning the word-form's association with either singular or plural. This is, however, not the case. There is no relationship between the length of the words and the accuracy of the number prediction (χ^2=10.4, df=6, p=0.11). This is expected since with more biphones one gets more association weight values, but these weights will not necessarily all add up to go in the same direction.

The predictive performance of our models may have been even further improved if gender would have been added as a separate cue. However, as discussed above, in comprehension, gender cannot be regarded as a cue in the same way as phonological cues. Rather, gender by itself is a grammatical feature that has to be accessed or computed from the input, and in German simplex nouns the mapping is largely non-transparent, even though recent studies have shown that gender can be accurately predicted on the basis of word usage as manifested in word embeddings (Seyboth & Domahs, 2023; Köpcke & Zubin, 1996; Nastase &

Popescu, 2009; Williams et al., 2019). In the case of the patient, deterioration of her lexical representations also led to impaired gender knowledge. Therefore we decided against including gender as an explicit cue. Yet, gender information was partly conveyed by the definite article, which was included as a cue.

With regard to the impaired lexicon, some of the differences of our findings from previous studies may be due to the modality addressed in both the patient data and our models: While many previous investigations focused on production (i.e. the making of a word form from a given meaning), in our case the opposite direction of processing was addressed, i.e. comprehension. This may explain different roles of cues (e.g. formal or semantic cues, gender) and different impact of factors like validity or perceptual salience. In a similar vein, in contrast to previous patient studies (e.g. Biedermann et al. 2018) we did not find a general advantage for singular – neither in the patient data nor in the modeling results. However, in the study by Biedermann et al. (2018) a production task (picture naming) was used. It may be that their observed singular advantage is specific to that modality.

In conclusion, it is possible to quite successfully predict the number of a given German noun based on the associations of its constituent biphones with plural and singular. Higher order generalizations may fall out of such an account automatically and need not be specified independently. Discriminative learning networks are able to model language systems, be they impaired or non-impaired.

References

Anderson, J. R. & C. J. Lebiere. 1998. *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Arndt-Lappe, S. & M. Ernestus. 2021. Morphology-phonology interaction. In V. Pirrelli, I. Plag & W. U. Dressler (eds.), *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*, 191–227. Berlin & New York: De Gruyter Mouton.

Arndt-Lappe, S., R. Schrecklinger & F. Tomaschek. 2022. Stratification effects without morphological strata, syllable counting effects without counts: Modelling English stress assignment with Naive Discriminative Learning. *Morphology* doi: https://doi.org/10.1007/s11525-022-09399-9.

Baayen, R. H., P. Milin, D. Filipović Durdević, P. Hendrix & M. Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438–481. doi: 10.1037/a0023851.

Baayen, R. H., R. Piepenbrock & L. Gulikers. 1996. Celex2. *Linguistic Data Consortium, Philadelphia* .

Behrens, H. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics* 47(2). 383–411. doi: https://doi.org/10.1515/LING.2009.014.

Beser, D. 2021. Falling through the gaps: Neural architectures as models of morphological rule learning URL `https://arxiv.org/abs/2105.03710`.

Biedermann, B., E. Beyersmann, C. Mason, F. Machleb, M. Moormann & A. Lorenz. 2018. Production of German -n plurals in aphasia: Effects of dominance and predictability. *Aphasiology* 32(9). 1087–1112. doi: 10.1080/02687038.2017.1387228.

Boswijk, V. & M. Coler. 2020. What is salience? *Open Linguistics* 6. 713–722. doi: 10.1146/annurev-linguistics-011619-030303.

Buch, A. 2011. *Linguistic spaces: Kernel-based models of natural language.* Tübingen: Universität Tübingen Phd dissertation.

Chuang, Y.-Y., K. Loo, J. P. Blevins & R. H. Baayen. 2020. Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning. In L. Körtvélyessy & P. Štekauer (eds.), *Complex Words: Advances in morphology*, 119–141. Cambridge: Cambridge University Press.

Daelemans, W. 2002. A comparison of analogical modeling of language to memory-based language processing. In R. Skousen, D. Lonsdale & D. B. Parkinson (eds.), *Analogical modeling: An exemplar-based approach to language*, 157–179. Amsterdam: John Benjamins.

Daelemans, W., J. Zavrel, K. van der Sloot & A. van den Bosch. 2007. *TiMBL: Tilburg Memory Based Learner, version 6.0, Reference Guide: LK Technical Report 04-02.* Tilburg: ILK.

Dankers, V., A. Langedijk, K. McCurdy, A. Williams & D. Hupkes. 2021. Generalising to German plural noun classes, from the perspective of a recurrent neural network. In *Proceedings of the 25th conference on computational natural language learning*, 94–108. URL `https://aclanthology.org/2021.conll-1.8`.

Danks, D. 2003. Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology* 47. 109–121.

Diessel, H. (ed.). 2019. *The grammar network.* Cambridge: Cambridge University Press.

Domahs, F., L. Bartha, A. Lochy, T. Benke & M. Delazer. 2006. Number words are special: Evidence from a case of primary progressive aphasia. *Journal of Neurolinguistics* 19(1). 1–37. doi: https://doi.org/10.1016/j.jneuroling.2005.07.001.

Domahs, F., L. Bartha-Doering, U. Domahs & M. Delazer. 2017. Wie muss ein "guter" deutscher Plural klingen? In N. Fuhrhop, R. Szczepaniak & K. Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 205–237. Berlin & Boston: De Gruyter Mouton. doi: https://doi.org/10.1515/9783110528978-008.

Eisenberg, P. & N. Fuhrhop. 2020. *Grundriss der deutschen Grammatik - Das Wort*. Stuttgart & Weimar: J.B. Metzler 5th edn.

Friedman, L. & M. Wall. 2005. Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression. *The American Statistician* 59(2). 127–136. URL `http://www.jstor.org/stable/27643645`.

Fábregas, A. & M. Penke. 2021. Word storage and computation. In V. Pirrelli, I. Plag & W. U. Dressler (eds.), *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*, 455–50582. Berlin & Boston: De Gruyter Mouton.

Grandon, B., M. Schlechtweg & E. Ruigendijk. 2023. Processing of noun plural marking in German-speaking children: an eye-tracking study. *Journal of Child Language* online.

Hahn, U. & R. C. Nakisa. 2000. German inflection: Single route or dual route? *Cognitive Psychology* 41(4). 313–360. doi: https://doi.org/10.1006/cogp.2000.0737.

Heitmeier, M., Y. Chuang, S. Axen & R. H. Baayen. 2024. Frequency-informed Linear Discriminative Learning. *Frontiers in Human Neuroscience* 17.

Heitmeier, M., Y.-Y. Chuang & R. H. Baayen. 2023. How trial-to-trial learning shapes mappings in the mental lexicon: Modelling Lexical Decision with Linear Discriminative Learning. *Cognitive Psychology* 146. 101598.

Hilpert, M. 2019. Higher-order schemas in morphology: What they are, how they work, and where to find them. *Word Structure* 12(3). 261–273. doi: https://doi.org/10.3366/word.2019.0149.

Hothorn, T. & A. Zeileis. 2015. partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research* 16(1). 3905–3909.

Janssen, U. & F. Domahs. 2008. Going on with optimised feet: Evidence for the interaction between segmental and metrical structure in phonological encoding from a case of primary progressive aphasia. *Aphasiology* 22(11). 1157–1175. doi: https://doi.org/10.1080/02687030701820436.

Kamin, L. J. 1969. Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (eds.), *Punishment and aversive behavior*, 276–296. New York: Appleton-Century-Crofts.

Köpcke, K.-M. 1988. Schemas in German plural formation. *Lingua* 74(4). 303–335.

Köpcke, K.-M. 1993. *Schemata bei der Pluralbildung im Deutschen: Versuch einer kognitiven Morphologie*, vol. 47 Studien zur deutschen Grammatik. Tübingen: G. Narr.

Köpcke, K.-M. 1998. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language* 25(2). 293–319. doi: https://doi.org/10.1017/S0305000998003407.

Köpcke, K.-M., S. Schimke & V. Wecker. 2021. Processing of German noun plurals: Evidence for first- and second-order schemata. *Word Structure* 14(1). 1–24. doi: 10.3366/word.2021.0173.

Köpcke, K.-M. & V. Wecker. 2017. Source- and product-oriented strategies in L2 acquisition of plural marking in German. *Morphology* 27(1). 77–103. URL `https://link.springer.com/article/10.1007/s11525-016-9292-5`.

Köpcke, K.-M. & D. Zubin. 1996. Prinzipien für die Genuszuweisung im Deutschen. Deutsch - typologisch, 473 – 491. Berlin [u.a.]: de Gruyter. doi: 10.1515/9783110622522-021. URL `https://nbn-resolving.org/urn:nbn:de:bsz:mh39-89446`.

Lieber, R. 2021. *Introducing morphology.* Cambridge: Cambridge University Press.

Luo, X. 2021. *JudiLing: An implementation for Discriminative Learning in Julia.* Tübingen Eberhard Karls University of Tübingen Master thesis. URL `https://github.com/MegamindHenry/JudiLing.jl/blob/master/thesis/thesis`$_JudiLing_{An_implementati}$

Luo, X., Y.-Y. Chuang & R. H. Baayen. 2023. JudiLing: an implementation in Julia of Linear Discriminative Learning algorithms for language model URL `https://https://github.com/MegamindHenry/JudiLing.jl`.

MacLeod, B. 2015. A critical evaluation of two approaches to defining perceptual salience. *Ampersand* 2. 83–92. doi: 10.1016/j. amper.2015.07.001.

Marcus, G. F., U. Brinkmann, H. Clahsen, R. Wiese & S. Pinker. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology* 29(3). 189–256. doi: https://doi.org/10.1006/cogp.1995.1015.

McCurdy, K., S. Goldwater & A. Lopez. 2020a. Inflecting when there's no majority: limitations of encoder-decoder neural networks as cognitive models for German plurals URL `https://arxiv.org/abs/2005.08826`.

McCurdy, K., A. Lopez & S. Goldwater. 2020b. Conditioning, but on which distribution? Grammatical gender in German plural inflection. In *Proceedings of the workshop on cognitive modeling and computational linguistics*, 59–65. URL `https://aclanthology.org/2020.cmcl-1.8`.

Nastase, V. & M. Popescu. 2009. What's in a name? In some languages, grammatical gender. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, 1368–1377. Singapore.

Nieder, J., Y.-Y. Chuang, R. van de Vijver & R. H. Baayen. 2022. A Discriminative Lexicon approach to word comprehension, production and processing: Maltese plurals. Accepted for publication. *Language* URL `https://psyarxiv.com/rkath/`.

Pearce, J. M. & M. E. Bouton. 2001. Theories of associative learning in animals. *Annual review of psychology* 52(1). 111–139. doi: https://doi.org/10.1146/annurev.psych.52.1.111.

Penke, M. & M. Krause. 2002. German noun plurals: A challenge to the Dual-Mechanism Model. *Brain and Language* 81(1-3). 303–311. doi: 10.1006/brln.2001.2526.

Penke, M., E. Wimmer, J. Hennies, M. Hess & M. Rothweiler. 2016. Inflectional morphology in German hearing-impaired children. *Logopedics Phoniatrics Vocology* 41(1). 9–26.

Pescuma, V. N., C. Zanini, D. Crepaldi & F. Franzon. 2021. Form and function: A Study on the distribution of the inflectional endings in Italian nouns and adjectives. *Frontiers in Psychology* 12. 720228. doi: https://doi.org/10.3389/fpsyg.2021.720228.

Pirrelli, V., C. Marzi, M. Ferro, F. A. Cardillo, H. R. Baayen & P. Milin. 2021. Psycho-computational modelling of the mental lexicon. In V. Pirrelli, I. Plag & W. U. Dressler (eds.), *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*, 23–82. Berlin & Boston: De Gruyter Mouton.

Plag, I. 2018. *Word-formation in English, 2nd edition.* Cambridge: Cambridge University Press.

Polišenská, D. 2010. *Dutch children's acquisition of verbal and adjectival inflection.* Amsterdam: Universiteit van Amsterdam Phd dissertation.

Ramscar, M., D. Yarlett, M. Dye, K. Denny & K. Thorpe. 2010. The Effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957. doi: 10.1111/j.1551-6709.2009.01092.x.

Ramscar, M. & D. Yarlett. 2007. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive science* 31(6). 927–960.

Rath, D., F. Domahs, K. Dressel, D. Claros-Salinas, E. Klein, K. Willmes & H. Krinzinger. 2015. Patterns of linguistic and numerical performance in aphasia. *Behavioral and Brain Functions* 11. 2. doi: 10.1186/s12993-014-0049-1.

Rescorla, R. A. 1988a. Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience* 11(1). 329–352. doi: https://doi.org/10.1146/annurev.ne.11.030188.001553.

Rescorla, R. A. 1988b. Pavlovian conditioning. It's not what you think it is. *American Psychologist* 43(3). 151–160. doi: 10.1037/0003-066X.43.3.151.

Rescorla, R. & A. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton-Century-Crofts.

Rosen, E. 2022. Modeling human-like morphological prediction. In *Proceedings of the Society for Computation in Linguistics 2022*, 133–142. URL `https://aclanthology.org/2022.scil-1.11`.

Schmid, H.-J. & F. Günther. 2016. Toward a unified socio-cognitive framework for salience in language. *Frontiers in Psychology* 7. 1110. doi: 10.3389/fpsyg.2016.01110.

Seyboth, M. & F. Domahs. 2023. Why do He She Disagree: The Role of Binary Morphological Features in Grammatical Gender Agreement in German. *Journal of Psycholinguistic Research* 52. 923–955. doi: 10.1007/s10936-022-09926-z.

Simoens, H., A. Housen & L. De Cuypere. 2017. The effect of perceptual salience on processing L2 inflectional morphology. In S. M. Gass, P. Spinner & J. Behney (eds.), *Salience in second language acquisition*, 107–130. New York: Routledge.

Tomaschek, F. & R. H. Baayen. 2017. Articulatory patterns of monomorphemic and dimorphemic homophonous words. In *FOR 2373 'Spoken Morphology' Colloquium Series (talk)*, .

Tomaschek, F., U. Domahs & F. Domahs. 2023. Modeling German word stress. *Glossa* 8. 1–28. doi: https://doi.org/10.16995/glossa.9015.

Tomaschek, F., I. Plag, M. Ernestus & R. H. Baayen. 2021. Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics* 57(1). 123–161. doi: https://doi.org/10.1017/S0022226719000203.

Tomaschek, F. & M. Ramscar. 2022. Understanding the phonetic characteristics of speech under uncertainty-implications of the representation of linguistic knowledge in learning and processing. *Frontiers in Psychology* 13. 754395. doi: 10.3389/fpsyg.2022.754395.

Twain, M. 1880. *A Tramp Abroad*. Hartford, Connecticut: American Publishing Company.

van de Vijver, R. & E. Uwambayinema. 2022. A word-based account of comprehension and production of Kinyarwanda nouns in the Discriminative Lexicon. *Linguistics Vanguard* URL `https://psyarxiv.com/78a6e/`.

Wiese, R. 2000. *The phonology of German*. Oxford: Oxford University Press.

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.

Williams, A., A. Drozdov & S. R. Bowman. 2019. Do latent tree learning models identify meaningful structure in sentences. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, 253–267. Florence, Italy: Association for Computational Linguistics. doi: 10.1162/tacl$_{a0}$0019.URL`https://www.aclweb.org/anthology/Q18-1019.pdf`.

Wulf, D. J. 2002. Applying Analogical Modeling to the German plural. In R. Skousen, D. Lonsdale & D. B. Parkinson (eds.), *Analogical modeling: An exemplar-based approach to language*, 109–122. Amsterdam: John Benjamins. doi: https://doi.org/10.1075/hcp.10.09wul.