Jennifer Keller and Ingo Plag

Discriminative learning of number interpretation of German pseudo-nouns

Abstract: The German nominal number system has long been a test case for theories of morphological organization and morphological processing. Recent research using discriminative learning networks (e.g. Plag et al., 2024, 2025) has challenged more traditional approaches such as the schema account (Köpcke and colleagues 1988; 1993; 2021). It has been shown, for instance, that discriminative learning models based on only biphones are able to predict not only whether a given word-form is singular or plural, but also the number decisions on real words by an aphasic patient.

The present study tests whether a discriminative learning model is also able to predict the number decisions on pseudo-words, as given in the data set from Köpcke et al. (2021). Those authors conducted a number decision task on German pseudo-words testing the assumptions of the schema account. The association weights from a Naïve Discriminative Learning (NDL) model of the present study turned out to be highly successful predictors of the choices of the participants in Köpcke et al.'s experiment, at the same level of accuracy as the schema-based regression model. Further explorations of the NDL model revealed that the strength of schema cues correlate with the strengths of the NDL associations.

Discriminative learning is thus able to account for crucial properties of the German nominal inflectional system and thus offers a new way of understanding the mapping of form and meaning in the German nominal system.

Keywords: number inflection, discriminative learning, form-meaning mapping

Jennifer Keller, Department of English language and linguistics, Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, e-mail: jennifer.keller@uni-duesseldorf.de
Ingo Plag, Department of English language and linguistics, Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, e-mail: ingo.plag@uni-duesseldorf.de

1 Introduction

There is an ongoing debate in morphological and psycholinguistic theory on how morphological structure is organized and processed in the mental lexicon. A particular prominent case in this debate is the highly complex German nominal number system. In this language, singulars are not marked morphologically, and plural is expressed in various ways. The most common indicators of plural ('cues') are illustrated in Table 1. The column 'Singular' illustrates the cues in non-plural contexts.

Tab. 1: Most common plural markers in German (cf. Köpcke 1993; Plag et al. 2024).

Cue for Plural	Singular/Plural Pair	Singular
-e	*Tisch* ('table') / *Tische* ('tables')	*Sage* ('myth')
-en	*Tür* ('door') / *Türen* ('doors')	*Boden* ('floor')
	Rose ('rose') / *Rosen* ('roses')	
-er	*Kind* ('child') / *Kinder* ('children')	*Kater* ('male cat')
-s	*Auto* ('car') / *Autos* ('cars')	*Fuchs* ('fox')
umlaut	*Vater* ('father') / *Väter* ('fathers')	*Bär* ('bear')
umlaut + *-e*	*Kuh* / *Kühe* ('cow')	*Tüte* ('bag')
umlaut + *-er*	*Wald* ('wood') / *Wälder* ('woods')	*Fächer* ('fan')
-∅	*Adler* ('eagle') / *Adler* ('eagles')	*Adler* ('eagle')
def. article	*der* (*Tisch*) / *die* (*Tische*)	*die* (*Tür*) / *die* (*Türen*)
	das (*Kind*) / *die* (*Kinder*)	

In this system, number is marked on the noun by a variety of suffixes (*-e, -en, -er, -s*), by umlaut, or a combination of suffix and umlaut. Zero marking is also found. While these markers are typically associated with plural formation, they are not exclusive to plurals. In fact, the forms can also appear in singular word-forms. For instance, the final schwa (or <e> in the spelling) signals plurality in *Tische* ('tables') but also occurs in the singular form of nouns, e.g. in *Sage* ('myth').

In addition to these formal cues, gender and the prosodic pattern of the noun have been proposed to have a say in the system (e.g. Wiese 2000, 2009). Syntactically, there is also number agreement with determiners and adjectives.

While the earlier debates were characterized by a discussion of competing processes ('words and rules'), more recent approaches can be grouped into two kinds: the usage-based schema account (Köpcke 1988, 1993; Köpcke et al. 2021) and a range of computational modeling approaches. Although both kinds of approach have been able to give valuable insights into the processing of number in German nouns, they suffer from certain drawbacks concerning proper implementation and the inclusion of the role of individual formal cues. More recently, discriminative learning, a theoretical approach, which is also computationally implemented (Baayen et al. 2011, 2019), has been introduced as a promising alternative. Plag et al. (2024, 2025), in particular, have demonstrated that discriminative learning models based on only biphones (instead of traditional units such as stress, suffixes and the like) are able to predict whether a given word-form is singular or plural. It has also been shown that the number decisions on given word-forms by an aphasic patient can be successfully predicted using a discriminative network.

Existing accounts mostly focus on the processing of existing words. However, an important litmus test for any model is the treatment of pseudo-words. It is generally assumed that inflection on words that are not stored in the speaker's mind, as in Berko-Gleason's classic *wug* test (Berko-Gleason 1958), can give crucial insight into the generative properties and capacities of the morphological system. Extant classical and deep learning computational models struggle with pseudo-words (e.g. Rosen 2022; Beser 2021), and there is only one study available that tests the schema account against pseudo-words (Köpcke et al. 2021).

Köpcke et al. (2021) conducted a number decision task on German pseudo-words testing the schema-based assumptions. The schema account was able to explain a large proportion of the participants' decisions, and thus provided evidence for the hypothesized strength of different formal cues. However, at a theoretical level, it is unclear how the differences in strength between the cues are derived, or learned by the speakers.

In the present study, we test, as an alternative to the schema account, a discriminative learning approach with pseudo-words (as against real words, as done in Plag et al. 2024, 2025), using the experimental data from Köpcke et al.'s (2021) study.[1] In particular, we compare the performance of our discriminative learning model with the predictions of the schema account. We explore whether the strength of a given cue may emerge as a result of discriminative learning.

Our results show that the association weights from a Naïve Discriminative Learning (NDL) model are highly successful in predicting the decisions of the participants in Köpcke et al.'s experiment, and do so at the same level

1 We are grateful to these authors for having provided us with their data set.

of accuracy as the schema-based regression model. Further explorations of the NDL model reveals that the strength of schema cues correlate with the strengths of the NDL associations. This suggests that the effects observed from the perspective of the schema account can be conceived as (at least partly) emerging from discriminative learning.

Discriminative learning is thus able to account for crucial properties of the German nominal inflectional system and thus offers a new way of understanding the mapping of form and meaning in the German nominal system.

The paper is structured as follows. In the next section, we provide an overview of the three approaches to German nominal number and outline the aims of the present study, including its research questions and objectives. Section 3 presents our method and analysis. The results are reported in Section 4 and afterwards discussed in Section 5.

2 Approaches to German nominal number

2.1 Schemas

The usage-based view of morphology holds that the lexical process and representation of words is influenced by the way speakers use language (cf. Bybee 2006). Speakers draw on phonological and semantic similarities among stored forms to determine the grammatical structure of a given word.

Many theoretical frameworks have been developed within this view, one of which is the schema account proposed by Köpcke (1988, 1993). In this account, speakers are assumed to make morphological decisions based on probabilistic generalizations derived from patterns in the lexicon. In general, research on morphological processing has shown that probabilistic generalizations can be based on different measures such as type and token frequency, relative frequency, phonological similarity, as well as entropy and surprisal (e.g. Baayen et al. 2007; Moscoso del Prado Martín et al. 2004). The schema account includes only some of these measures, frequency-based measures in particular, to formulate such generalizations.

Köpcke et al. (2021, 5) define a generalization, or 'schema', as "an abstraction of many known word forms [...] that belong to a specific grammatical category". One such grammatical category is, for example, the plural. Nouns with similar endings, phonological shapes, or gender often share the same morphological patterns of a grammatical category, which then constitute a schema. A schema can be, for instance, plural patterns like *Xe* as in *Regale, Hunde, Tische* ('cupboards, dogs, tables').

Schemas can be further divided into two types: first-order and second-order schemas. Plural patterns like *Xe* represent first-order schemas, which refer to the structure of word-forms based solely on their surface form. By contrast, second-order schemas relate two first-order schemas, thereby describing the relationship between forms within a paradigm (Hilpert 2019; Köpcke 1988). For example, relating the first-order schema *die Xe* as in *die Katze* ('the cat') with the first-order schema *die Xen* as in *die Katzen* ('the cats') results in a second-order schema that reflects the singular-plural relationship.

As discussed in Section 1, one problem of the German nominal number system is that formal cues for plural may also occur in singular word-forms. As schemas work in a probabilistic manner, they do not determine categorically whether a given word-form is singular or plural. Instead, they vary in how reliably they signal a certain number. In other words, schemas differ in their strength as indicators of singularity or plurality. According to Köpcke (1988), the strength of an individual cue to signal plural or singular is determined by three distributional and perceptual factors: salience, (type) frequency, and validity.

The 'salience' of a cue refers to how perceivable the different plural markers are in the acoustic signal. This is based on the fact that some cues are easier to perceive than others (Köpcke 1998). Two of Slobin's (1973) operating principles suggest that segmental cues, or cues that appear at the beginning or the end of a word, such as the suffixes *-(e)n, -s, -e* and *-er*, are generally more salient than cues that are non-segmental or appear word-medially, such as umlaut (recent empirical evidence is shown in Grandon et al. 2024). Cue salience also varies with complexity: cues with multiple segments are more salient than those with just one segment. In a similar vein, cues with a full vowel are better perceived than cues with a reduced vowel, followed by cues with no vowel, followed by cues with zero marking (Polišenská 2010).

The 'type frequency' of a cue denotes the number of nouns that share a particular plural marker. According to Köpcke (1988), the most frequent plural marker is *-(e)n* followed by *-e*. The plural markers *-s, -er* and umlaut occur with low frequency.

Finally, the 'validity' of a cue indicates how often (across the lexicon) a particular form (e.g. final schwa) is used to mark one grammatical function, such as plural, compared to other grammatical functions, such as singular. The validity of a cue is thus a type-based measure of relative frequency.

Recent work by Köpcke et al. (2021) emphasizes that this factor is crucial in determining how reliable a certain cue is in signaling plurality. A cue that occurs very frequently in plural word-forms but rarely in singular word-forms serves as a strong indicator of plural meaning. The validity of a cue to signal plurality increases as the difference between its plural and singular type frequencies becomes larger. According to Köpcke (1988), cue validity is high for the suffixes *-(e)n* and *-s*. The suffixes *-e* and *-er* have low cue validity. Cue validity for the umlaut is generally relatively low.

The idea of cue validity is consistent with ideas about the role of type frequency and token frequency in usage-based frameworks which assume that forms with high type frequency are more likely to be represented and processed in schemata, while forms with high token frequency tend to be represented and processed as whole-word items (Behrens 2009; Diessel 2019).

The three factors, i.e. salience, (type) frequency, and validity, and their influence on cue strength are summarized by Köpcke (1988, 1993) in the table given below (Table 2). The plus sign indicates that a particular factor is valid for the cue listed. As can be seen, suffix *-(e)n*, for instance, is assumed to be a strong plural marker due to its high type frequency, salience and cue validity. In contrast, umlaut is predicted to be a weak plural marker since its cue validity, salience and type frequency are low.

Tab. 2: Cue strength of plural markers (based on Köpcke 1988, 315).

Marker	Salience	(Type) frequency	Cue validity
-(e)n	+	+	+
-s	+	−	+
-e	+	+/−	−
-er	+	−	−
umlaut	−	−	+/−

In more recent work, Köpcke et al. (2021) summarize the cue strength of plural markers along a continuum, as shown in Figure 1. According to this scale, *-en* is predicted to be the strongest plural cue, followed by *-s*, then the combination of umlaut and *-e*, and then *-e* itself. Monosyllabic stem is assumed to be the weakest cue for plural.

Previous work on the German nominal number system suggests that number schemas have played a role in diachronic change and influence how speakers process nominal number marking German (e.g. Köpcke 1988, 1993). Schemas have also been shown to affect the use of grammatical number in

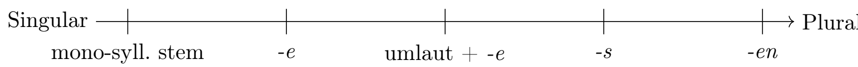

Fig. 1: Cue strength of plural markers on a continuum (based on Köpcke et al. 2021, 8)

both first and second language learning (Köpcke 1998; Köpcke & Wecker 2017), as well as the processing of number in individuals suffering from an impaired language system (Domahs et al. 2017).

An important issue for any approach to understanding inflectional systems is the treatment of novel forms by the speakers. Investigating the way speakers deal with unfamiliar words is crucial as it provides direct evidence of the productivity of morphological patterns beyond memorized forms (e.g. Berko-Gleason 1958). Since speakers cannot access novel forms in their lexical memory, speakers may use productive rules (such as a 'default rule'), or the probabilistic nature of morphological generalizations and the interplay of competing schemas (Köpcke 1988, 1993; Marcus et al. 1995).

Over the past decades, experimental studies have examined whether speakers apply known inflectional patterns to novel forms by employing pseudo-words, i.e. word-forms that are inexistent in the language but conform to the language's phonological inventory and phonotactics. Berko-Gleason's (1958) classic *wug* test has demonstrated that both children and adults readily apply inflectional patterns as, for instance, the English plural suffix -*s* to novel nouns (e.g. *wug* – *wugs*). With respect to German nominal number, Köpcke et al. (2021) have shown that speakers assign number to pseudo-words in ways comparable to real words (e.g. by assigning plural to novel forms ending in -*s*, such as *Knolks*). These findings suggest that speakers draw on their knowledge of word shapes (or schemas) as derived from existing words, and apply them as cues to determine the number of novel forms.

Although the schema approach has been able to account for many aspects of number processing in German nouns, the notion of schema faces several challenges. First, previous studies have mostly relied on illustrative examples or anecdotal evidence for the reality of the schema. They were limited in a number of aspects: they incorporated a small set of words (both in corpus and psycholinguistic studies), included relatively few participants, and, importantly, encountered only a small number of formal cues. Second, prior work on the schema approach has not provided any proper formal implementation of how schemas are related to the processing of words. Third, and more critically, the predictions on the relative importance of individual cues within the schema account remain vague, not properly quantified, or are even lacking altogether. While Köpcke (1988) define cue strength in terms of salience, (type) frequency, and validity, it is unclear how these factors interact and whether some exert a greater influence on cue strength than others. The minuses and pluses in Table 2 are unclear in their specific quantitative effects. In other words, the relative weighting of these factors in determining the strength of a given cue remains unspecified.

Since the relative weighting of the cue strength factors salience, (type) frequency, and validity is not quantified, it is unclear whether the strength differences between the cues are the same. According to Figure 1, the assumptions of the schema account seem to imply equal intervals between cue strengths. There is also no threshold for singularity vs. plurality, which means that the approach provides no criteria for determining when a cue qualifies as a valid singular or plural marker.

Finally, there are two further issues concerning plural schemata: the first is gender, the second stress. It has frequently been shown that gender is also to some extent predictive of plural allomorphy by constraining allomorph selection: masculine and neuter nouns form their plural predominantly with *-e*, while feminine nouns prefer *-(e)n*. Particular phonotactic patterns may impose further preferences, such as the predominace of zero plurals with masculine and neuter nouns ending in a closed schwa-syllable (*-el, -er, -en*). In the experiment whose data we analyze in the present paper, gender was not considered as a relevant variable by Köpcke and colleagues (see Köpcke et al. 2021, 13-15). We also did not take gender into account.

The second problem concerns the role of stress patterns. There is sufficient evidence that typical plural forms prosodically can be characterized as left-strong binary feet, i.e. trochees (e.g. Wiese 2000, 2009). Although this aspect is not incorporated in Köpcke's (1988; 1993) schema account, it has been shown to be relevant (see also Plag et al. 2024).

2.2 Computational approaches

The second type of current approaches to accounting for the German nominal number system involves computational models. Traditional computational models have been shown to be valuable tools for systematically implementing proposed theoretical concepts and mechanisms, both for predicting and processing number. Various implementations have been proposed, employing a range of statistical methods and model architectures inspired by (psycho-)linguistic theory, including a model based on declarative memory and procedural rules (Anderson & Lebiere 1998), a dual-route model (Hahn & Nakisa

2000), and analogical models (Daelemans 2002; Wulf 2002; Daelemans et al. 2007; Rosen 2022; Buch 2011).

Deep learning models were also introduced in the context of number processing in German nouns, and these models were even able to outperform the previously mentioned approaches (e.g. McCurdy et al. 2020a,b; Beser 2021; Dankers et al. 2021). Despite their strong performance, however, deep learning models also come with specific limitations. First, their architectures are in many cases not transparent, as they often include, for example, multiple hidden layers and non-linear mapping functions, which makes linguistic interpretation very difficult. Second, deep learning models are often not based on a cognitively grounded learning theory and frequently diverge from observable human behavior.

Importantly, both the more traditional and the deep learning models show weaknesses when applied to pseudo-words. Rosen (2022), for example, developed a similarity-based analogical model, in which the plural form for a new lexeme is predicted from its singular by identifying a phonologically similar singular with a known plural and transferring its plural class to the new lexeme. While this model was able to account for existing words, it performed poorly on pseudo-words. Beser (2021) encountered similar limitations with transformer and recurrent neural networks, which predicted the plural forms of pseudo-words from character sequences of existing singular-plural noun pairs. Although their model was designed to generate plural forms of pseudo-words, it was not able to predict novel words accurately.

More recently, discriminative learning has been proposed as an alternative computational model to account for nominal number in German (Heitmeier et al. 2021; Plag et al. 2024, 2025). This new approach addresses several of the limitations associated with deep learning models as it incorporates a simple two-layer architecture that is accessible to linguistic interpretation. Since discriminative learning is built on a well-established learning theory (e.g. Rescorla & Wagner 1972; Rescorla 1988b), it offers a cognitively more plausible approach compared to deep learning systems. Furthermore, recent work in this domain has demonstrated that discriminative learning can successfully model pseudo-word behavior in an inflectional system. Nieder et al. (2023), for instance, used a discriminative learning approach to model plural class assignment in Maltese pseudo-words, comparing activation-based model predictions to speakers' behaviour in a wug test. While the present study addresses a different empirical domain, the modelling logic is similar in that plural class predictions are derived from learned cue–outcome associations and evaluated against speaker data.

2.3 Discriminative learning

Discriminative learning is a well-established theory of associative learning mechanisms in cognitive psychology (e.g. Rescorla 1988b; Pearce & Bouton 2001). These associative learning mechanisms have been demonstrated to successfully model a variety of learning phenomena in both human and animals (e.g. Kamin 1969; Ramscar et al. 2010). In recent years, discriminative learning has gained increasing attention in linguistics. A growing body of empirical research, particularly in the field of morphology, has demonstrated that discriminative learning provides a promising framework for modeling a wide range of linguistic processes and effects (e.g. Arndt-Lappe et al. 2022; Baayen et al. 2011, 2018, 2019; Blevins 2016; Chuang et al. 2020; Chuang & Baayen 2021; Chuang et al. 2023, 2021; Heitmeier et al. 2021; Plag et al. 2022, 2024, 2025; Ramscar & Yarlett 2007; Ramscar et al. 2010; Saito et al. 2020; Schmitz et al. 2021; Stein & Plag 2021; Tomaschek et al. 2021, 2023; Nieder et al. 2023; Van de Vijver et al. 2024).

Discriminative learning theory assumes that learning consists of building associations between events, in the case of language learning between language-related events. These associations may form the basis for building abstract representations of the real world. The associations between events, so-called ‘cues’ and ‘outcomes’, are updated with each learning event. The weights of these associations are increased whenever a particular cue and a particular outcome co-occur. Conversely, if that cue occurs in the absence of that outcome, the association between cue and outcome decreases. Figure 2 illustrates such a network. The two word forms *Hund* (‘dog’) and *Hunde* (‘dogs’) act as cues and the numbers singular and plural as outcome. The thickness of the lines represents the strength of the associations. In this figure, plural is more strongly associated with the form *Hunde* than with the form *Hund*.

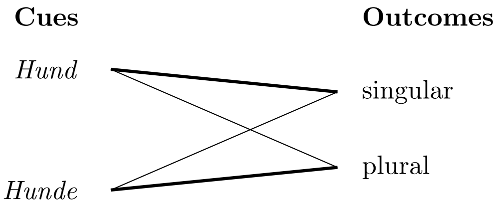

Fig. 2: Network including *Hund* (‘dog’) and *Hunde* (‘dogs’) as cues and number as outcome.

One particular implementation of discriminative learning theory is ‘Naïve Discriminative Learning’ (NDL) (Baayen et al. 2011). In an NDL model, the strength of the associations between cues and outcomes are calculated using the Rescorla-Wagner equations (Rescorla & Wagner 1972; Rescorla 1988a,b). After a certain number of learning events, the association weights reach a stable final state which indicates how strongly a given cue activates a corresponding outcome. An alternative method for directly calculating these final weights are Danks’ equilibrium equations (Danks 2003).

In the context of German nouns, there have been three recent studies which applied discriminative learning to predict the number of real words. One of these is Plag et al. (2024), who addressed the problem of schema cues, which is of particular relevance to the present paper. Plag et al. (2024) tested an NDL model based on cues as proposed by the schema account, and compared this model with a model that included only biphones as cues. Plag et al. (2024) demonstrated that a model using only biphones as cues was better in predicting the number of a given word-form than a schema-based model. Plag et al. (2024) also examined whether an NDL model is able to predict the response patterns of a lexically impaired patient in a number decision task. The model was trained on the patient’s responses with biphones as cues only, and was able to successfully predict the patient’s responses, suggesting that the model can capture even relevant aspects of impaired processing.

The other two discriminative learning studies are Heitmeier et al. (2021) and Plag et al. (2025). Heitmeier et al. (2021) developed a linear discriminative learning model (LDL, Baayen et al. 2019) which is able to produce and comprehend plural forms based on the mapping of phonological and semantic information. These authors also implemented a model to investigate pseudo-words, and that model’s performance closely resembled that of native German speakers. Plag et al. (2025) investigated the role of semantics in German nominal number marking using discriminative learning and found that, in addition to phonological information, the semantic properties of the nouns are also influential in distinguishing singular forms from plural forms.

In the present study, we investigate the role of phonological cues as proposed by the schema account in distinguishing singular and plural in pseudo-words, building on the work of Plag et al. (2024) on real words. The data set we use comes from Köpcke et al. (2021), who conducted a number decision task on German pseudo-nouns in order to test the assumptions of the schema account. We implement an NDL model (instead of an LDL model), for mainly two reasons. First, we are dealing with pseudo-words, which, according to the tacit assumptions by Köpcke et al. (2021), have no meaning, which implies that we are faced with a pairing of only form and number. Second, NDL is particular well-suited to model binary outcomes.

Our study has three aims. First, we test whether an NDL model based on bigrams can predict the participants' decisions in the number decision task. Second, we compare the predictions of our NDL models with the predictions of the schema account. Third, we quantify the relative strength of individual cues within the NDL framework, and explore whether the assumed strength of schema cues may be explained as effects of discriminative learning.

3 Method and analysis

We computed two types of models using the data set from Köpcke et al. (2021): a baseline regression model and NDL-based models. We aim to examine the properties of the two types of models by comparing a baseline model that relies on schema cues with NDL-based models that do not use such cues. Our first goal is to investigate whether the NDL-based models can successfully predict the decisions of participants in a number decision task without using any schema-based information.

For the baseline model, we implemented a regression model similar to that of Köpcke et al. (2021), which was based on the phonological cues from the schema account. For the discriminative learning approach, we implemented NDL-based networks using bigrams. Details on each model are provided as we go along. All models were implemented in R (R Core Team 2019) using the RStudio environment (Posit team 2025).

3.1 Data set

Köpcke et al. (2021) investigated the number interpretation of German pseudo-nouns under the assumptions of the schema account. Their findings showed that participants assigned number to novel word-forms according to the predicted cue strength of specific German plural markers. Specifically, *-en* and *-s* yielded the highest percentage of plural decisions, followed by the combination of umlaut and *-e*, then *-e* on its own, and finally monosyllabic stem (see again Figure 1).

In two experiments, Köpcke et al. (2021) tested 300 stimuli consisting of 60 monosyllabic nonce stems (e.g. *Knolk, Fump, Zauk, Broff, Kurm*), which were shown either as bare stems or in combination with one of the suffixes

-en, -s, -e, or umlaut together with *-e*. Participants completed two number decision tasks: in the first, pseudo-nouns were preceded by the article *die*; in the second, they were presented in isolation. For the present study, we used the participants' number decisions from the second number decision task, which involved the presentation of only the word-forms, without article. The reason for choosing the data set with the nouns presented in isolation is that the data set with the nouns preceded by the feminine/plural article *die* comes with an additional confound. This confound is the fact that the article *die* can be interpreted as either plural or feminine singular. Since gender also has a say in number assignment, it is unclear how the participants made use of this additional information in the stimuli. As a result, it is not surprising that Köpcke et al. (2021) struggle with the interpretation of the results of their experiment 1. The data set from the second number decision task, by contrast, does not include this confound, which makes the results more straightforward to interpret. The data set from experiment 2 comprised 2160 responses. In Example (1) we list the first 10 stems that were used in the experiment for illustration (see Köpcke et al. 2021, Appendix I):

(1) Wont, Blund, Zurf, Drolp, Grant, Blomp, Funt, Spruhn, Plon, Bold

3.2 Baseline model

As mentioned previously, we implemented a regression model very similar to that devised by Köpcke et al. (2021) for our baseline model, but with a more complex random effect structure. Using the `glmer` function from the `lme4` package (Bates et al. 2015), we fitted a logistic mixed-effects regression model to the observed data to create a model based on the schema cues. All variables listed in Table 3 entered our baseline model. In contrast to Köpcke et al. (2021), who in their random effect structure only included subject and item a random intercepts, we also added LEMMA as a categorical random effect variable to account for possible effects of the stem. The LEMMA of each word-form was identified by using its respective bare stem. Furthermore, we also tested random slopes for SCHEMACUE, ITEM, SUBJECT and LEMMA to find the optimal random effect structure. [2]

2 We renamed some of the variables for clearer understanding.

Tab. 3: Summary of variables (cf. Köpcke et al. 2021)

Variables	Levels
numberDecision	singular: 1085, plural: 1075
schemaCue	-en: 432, -s: 432, -e: 432, umlaut + -e: 432, stem: 432
item	300
subject	36
lemma	60

3.3 NDL-based models

Our aim was to implement NDL-based models that predict the association weights of pseudo-words with either number to determine whether a given pseudo-word is more likely to be classified as singular or plural. These pseudo-word associations are then used to predict the participants' decisions in the experiment. We implemented two kinds of model, which differed in the kinds of cues. In one model we used orthographical information and in the other phonological information. We did so based on the following considerations. The experiment was carried out in writing. Participants saw orthographic representations, and no audio signal was involved. This may suggest that it would be sufficient to implement a model using solely orthographical information. However, we decided to also implement a model with phonological cues as it is widely assumed that the two modalities are not independent from each other (e.g. Taft & Hambly 1985; Dixon et al. 2002; Brysbaert 2022). The orthography-based NDL model used bigraphs as cues, while the phonology-based NDL model used biphones as cues. Both NDL models included numberDecision as the outcome.

In order to obtain association weights of pseudo-words with singular and plural ('activations'), we implemented the following workflow. We first trained a network on real words using the CELEX lexical database (Baayen et al. 1996) to obtain association weights between bigraphs/biphones (cues) and singular/plural (outcomes). In the second step, we used the association weights of the individual bigrams from the real word network to compute the association weights of the pseudo-words with either number by adding up the association weights of the constituent bigrams for each pseudo-word. In the final step, we used these associations to predict the participants' decisions in the experiment. In the following, we will explain each step in more detail.

In the first step of our workflow, we trained a network on real words using mono-morphemic words in their singular and plural forms from CELEX (N= 7842) to obtain association weights between individual bigrams (i.e. either bigraphs or biphones) and the number of the word-form they were part of. Figure 3 illustrates an example of a network for the word *Puppe* (‘doll’), shown separately for the bigraph-based model (a) and the biphone-based model (b). For the bigraph-based model, we converted all initial uppercase letters, which is the common spelling in German nouns, to lowercase. Hash marks indicate word boundaries.

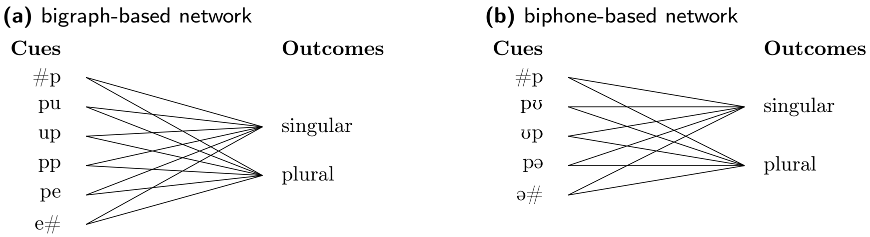

Fig. 3: NDL network for the word *Puppe* ('doll') using bigraphs as cues (left) and biphones as cues (right). Hash marks indicate word boundaries.

As explained in Section 2.3, the co-occurrence of a given cue and its paired outcome strengthens the association weight between the two, while the presence of a cue without that outcome weakens the association weight. With reference to Figure 3, the association weight between, for instance, the bigraph *pu* and the number `singular` is strengthened when they co-occur, and simultaneously weakened for the association between the bigraph *pu* and the number specification `plural`. To calculate the association weights between cues and outcomes, we applied the Danks' equilibrium equations (Danks 2003) using the `estimateWeights` and `estimateActivations` functions from the `ndl` package (Arppe et al. 2018). The computation also incorporated the frequency of each (real) word, allowing the equilibrium equations to simulate learning as if cue-outcome pairings had occurred according to their frequency. The information on frequency was taken from CELEX.

In the second step of our workflow, we used the plural and singular activations from the real word network to compute the association weights of the pseudo-words with either number. For this, we summed up the activations from real words for either number of each bigraph or biphone within a given pseudo-word to obtain its overall plural or singular activation. To give an example, Table 4 shows the singular and plural activations for the pseudo-word *Purte* in the bigraph- and biphone-based networks.

Tab. 4: Singular and plural activations for the pseudo-word *Purte* in the bigraph- and biphone-based networks.

bigraph-based							
singular	#p	pu	ur	rt	te	e#	Sum
	0.466	−0.137	−0.019	0.199	−0.360	0.412	0.561
plural	#p	pu	ur	rt	te	e#	Sum
	0.034	0.137	0.019	−0.199	0.360	0.088	0.439
biphone-based							
singular	#p	pʊ	ʊr	rt	tə	ə#	Sum
	0.999	−0.820	−0.328	0.599	−1.097	1.127	0.480
plural	#p	pʊ	ʊr	rt	tə	ə#	Sum
	−0.553	0.820	0.328	−0.599	1.097	−0.573	0.520

The total plural activation of *Purte* was calculated by adding the plural association weights of each of its bigraphs or biphones, as derived from the real-word network trained on CELEX. Analogously, we can calculate the singular activations. Based on these activations we can assume that, if a categorical decision is required, *Purte* is assigned singular in the bigraph-based model, and plural in the biphone-based model.

In the third and final step of our workflow, we calculated the difference between the plural activation and the singular activation for each word-form, and used this 'plural activation difference' to predict the participants' decisions in experiment. To arrive at these predictions, we fitted logistic mixed-effects regression models with the pseudo-word activation differences as predictor of interest and the number decisions of the participants as dependent variable, using the `glmer` function.

All data sets and analysis scripts are available at https://osf.io/9j7cq/overview?view_only=c1e0d080eb6b4598985bc51c1190e23d.

4 Results

4.1 Baseline model: Schema cues

We included NUMBERDECISION as the dependent variable and SCHEMACUE as the predictor of interest. The random effect structure was specified as follows. We tested random effects for SUBJECT, ITEM and LEMMA by devising first models that included only one effect, specified as a random intercept and a random slope. We used the Akaike Information Criterion (AIC) to asses the model fit and kept the random slope only if the AIC-value decreased in the presence of the random slope. Only the random slope for SUBJECT remained.

We then tested all combinations of the three intercepts (always including the random intercept and slope for SUBJECT), again using the AIC to test whether a random effect was justified. The final model had the random intercept and slope for SUBJECT and the random intercept for ITEM as the optimal random effect structure.

The partial effect of SCHEMACUE in the baseline model is shown in Figure 4. The different schema cues are presented on the x-axis. The y-axis gives the predicted probability of the participants' plural decisions in the experiment with higher values indicating a stronger tendency toward plural.

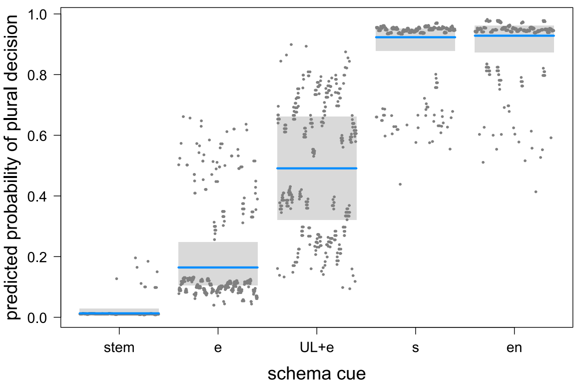

Fig. 4: Predicted probabilities (blue lines) of the participants' plural decisions in the experiment for each schema cue as predicted by the baseline model. Grey areas indicate 95% confidence intervals and dots represent the residuals. UL = umlaut.

The baseline model predicts that *-en* and *-s* have the highest probability of plural decisions in comparison with the other schema cues. The combination of umlaut and final *-e* (umlaut + *-e*) is predicted to have a probability of 50.1 percent, meaning it has no bias toward either plural or singular. Monosyllabic stem and *-s* are predicted to have low probabilities of plural decisions, with monosyllabic stem having the lowest. Hence, monosyllabic stems are predicted to have the strongest tendency toward a singular interpretation, followed by *-e*.

To perform pairwise comparisons for SCHEMACUE, we used the Tukey's Honest Significant Difference (HSD) method to perform post-hoc pairwise comparisons.[3] Using the `glht` function from the `multcomp package` (Hothorn et al. 2008), we found that all schema cues differed significantly in their probability of plural decisions ($p<0.001$), apart from the contrast between *-en* and *-s* ($p>0.1$). Overall, the results of our baseline model closely align with those reported in Köpcke et al. (2021).

While the three schema cues on the left behave as expected, *-s* and *-en* behave partly unexpected. Quite expectedly, they come out as the strongest cues, but there is no difference between them, contra to what the schema account would have predicted. *-s* overperforms, in that the participants have a stronger tendency towards plural decisions with *-s* final words than the schema approach predicted.

4.2 NDL-based models: Activations and plural decisions

To investigate whether the NDL associations can predict the decisions elicited from the participants in the experiment, we created logistic mixed-effects models that estimated the relationship between the pseudo-word associations and the observed plural decisions. The activation difference entered our logistic mixed-effects models as the predictor of interest, and NUMBERDECISION was included as the dependent variable. In the full model, we also added random intercepts and slopes for SUBJECT, ITEM, and LEMMA.

To determine the best-fitting random-effects structure, we followed the same procedure used for the baseline model. As the AIC-value decreased in the presence of the random slopes for each of the random effects SUBJECT, ITEM, and LEMMA, we first kept the model with random slopes for all three random effects. We then tested all combinations of the three random intercepts and random slopes, again using the AIC to test whether a random effect was justified. The model with the lowest AIC value included SUBJECT and ITEM as random intercepts and random slopes. This was the case for both the bigraph-based model and the biphone-based model.

3 Köpcke et al. (2021) applied a Bonferroni correction. The Bonferroni method is considered rather conservative: while it effectively reduces false positives (i.e. detecting a difference when non exists, Type I errors), it increases the likelihood of false negatives (i.e. not detecting a difference when one exists, Type II errors). Thus, one might miss real effects (see, for example, García-Pérez 2023 for discussion). To avoid this issue, compared to the Bonferroni method, Tukey's HSD method offers a more balanced alternative, which controls over both Type I and Type II errors while preserving greater statistical power.

Figure 5 presents the results of our logistic mixed-effects models. The left panel shows the partial effect of the pseudo-word activation difference derived from the NDL model using bigraphs. The right panel shows the analogous effect for the model that uses the NDL activations based on biphones. In both panels, the x-axis represents the activation difference with higher values indicating stronger plural activations. The y-axis gives the predicted probability of the participants' plural decisions in the experiment with higher values indicating a stronger tendency toward plural and lower values a stronger tendency toward singular.

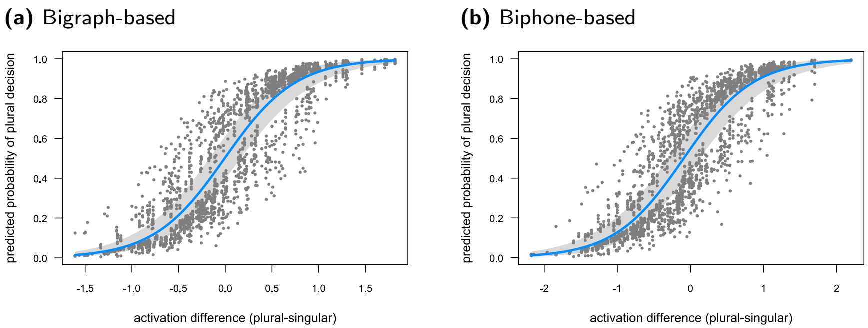

Fig. 5: Partial effect of plural activations (as against singular activations) on the predicted probability of plural decision in the mixed-effects regression models including residuals. Grey areas indicate 95% confidence intervals.

For both models, we can see that there is a strong correlation between the activation difference and the predicted probability of plural decision. That is, the higher the activation difference, the higher the predicted probability that participants choose the plural form in the experiment. This means that the activations derived from the NDL-based models are effective predictors of participants' plural decisions.

4.3 Comparison of baseline model and NDL-based models

To compare the model fits of the baseline and NDL-based models, we first assessed how well each model discriminated between the two outcome classes (`plural` versus `singular`) using the `Cstat` function from the `DescTools` package (Signorell 2024). All models were well able to discriminate plural against singular. The performance (measured in terms of the concordance index C)

of the bigraph-based and the biphone-based NDL model (C=0.95 for both) was similiar to the performance of the baseline model (C=0.95).

For further model fit comparison, we used the `confusionMatrix` function from the `caret` package (Kuhn 2008) to compute cross-tabulations of observed classes versus predicted classes (using a threshold of 0.5 for the categorization of the activation difference into plural vs. singular). Table 5 summarizes the main results from the confusion matrices for each model. We use the $F1$ score as a balanced measure of prediction accuracy.[4]

Tab. 5: Overview of the accuracies of the three regression models. Figures are rounded to two digits.

Type of model	Accuracy ($F1$)	Singular prediction	Plural prediction
Baseline model	0.88	0.88	0.89
Bigraph-based NDL model	0.88	0.88	0.88
Biphone-based NDL model	0.88	0.87	0.89

The predictions of the bigraph-based and biphone-based NDL model are very similar in accuracy. While both NDL-based models reached a high accuracy, the bigraph-based NDL model was slightly better in predicting singular correctly and the biphone-based NDL model in predicting plural decisions correctly. NDL-based models perform at the same level as the baseline model.

In sum, the results demonstrate that NDL-based models have a very similar fit to the participants' number decisions as the baseline regression model that uses schema-based cues. In the following section, we will explore how the NDL model relates to the schema-based model to investigate the question of how it is possible that two models that use very different representations can arrive at very similar predictions.

4 The F1 score is the harmonic mean of two measures, precision and recall. 'Recall' is the number of items for which a model correctly predicts a certain outcome divided by the number of items which have actually have that outcome (i.e. the ratio of true positive cases to all cases of this class). It is thus a measure of how well the model is able to find a certain outcome. 'Precision' is the number of items for which the model correctly predicts a given outcome divided by the number of all items for which the model predicts that outcome (i.e. the ratio of true positives and all positives). Precision thus tells us how well aimed the model is in its predictions.

4.4 Relating activations to schema cues

In this section, we will inspect the NDL model in more detail to see whether the NDL activations can be related to the schema cues. In Figure 6, we take a closer look at the activation difference of the individual items in the experiment in order to see how the five different classes of stimuli (representing the five kinds of schema cues) are treated by the NDL model. We use only the bigraph-based NDL model here for reasons of space, and because the biphone model is very similar in its results (see supplementary material). In Figure 6, we plot the activation difference of each item against the proportion of plural decisions found in the experimental data.

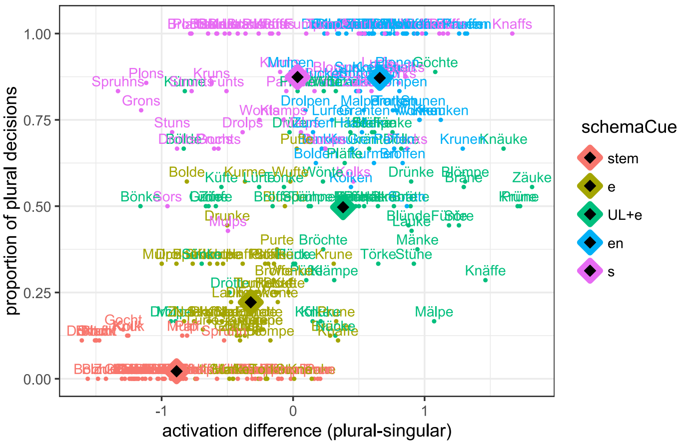

Fig. 6: Correlation between the plural-singular activation difference of the bigraph model (on the x-axis) and the proportion of plural decisions for each word-form (on the y-axis), with cues for plurality color-coded: red = stem, light green = *-e*, dark green = umlaut + *-e*, blue = *-en*, pink = *-s*. The diamonds represent the means of the activations differences in the five categories.

We can see that word-forms exhibiting the same schema cue tend to form clusters according to their activation differences and proportions of plural decisions. Word-forms with monosyllabic stems and *-e*, which are assumed in the schema account to be rather associated with singular, form a cluster at the lower end of both the activation difference scale and the proportion of plural decisions scale. Word-forms ending in *-en*, the supposedly strongest cue for plurality, cluster at the higher end of both the activation difference and proportion of plural decisions scales. For words with umlaut + *-e*, a cue that is assumed to have no tendency toward either plural or singular, there is a rather large cluster stretching all across both scales, with a positive correlation of activation difference and the proportion of plural decisions.

Word-forms including *-s*, which in the schema model is assumed to be the second strongest cue for plurality, do not cluster in line with their proportion of plural decisions. Instead, they are scattered across the activation difference scale, showing no clear correlation with the proportion of participants' plural decisions.

A look at the means of the by-word activation differences supports our eyeballing analysis. The diamonds in Figure 6 represent the means of the activation differences of five categories. One can see that, apart from *-s*, the increase of the means on the x-axis goes together with a rise in the proportion of plural decisions. This is in accordance with the scale of plural schema strength as given in Figure 1. The means are listed in Table 6.

Tab. 6: Means of activation differences by schema cue.

Schema cue	Mean of activation difference
stem	−0.89
-e	−0.32
umlaut + *-e*	0.38
-s	0.03
-en	0.66

These findings indicate that, via the by-word activations, the NDL-based model is able to pick up what is conceived as morphological structure in the schema-based account, although the NDL model is given no explicit morphological information. This holds for all cues but final *-s*, which, going by the activation difference, should not go together with the pronounced tendency for plural decisions in the experiment.

The implementation of NDL-based models also allows us to quantify the strength of the association between the bigrams representing the different cues on the one hand, and the two number specifications on the other. To derive the different bigram-based strengths of schema cues from an NDL-based model, we first extracted the real word activation differences for the bigrams of the schema cues *-e* and *-s*, which correspond to a single bigraph (i.e. *e#* and *s#*) or biphone (i.e. ə# and s#). For the schema cue *-en*, we calculated the activation differences by summing up the activation differences of the bigraphs *en* and *n#*, and, for the biphone-based model, the activation

(a) Biphone-based model

(b) Bigraph-based model

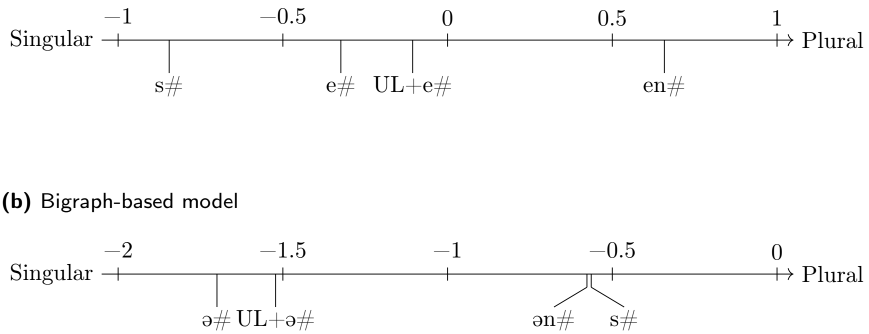

Fig. 7: Activation differences for the schema cues pitted against the schema continuum. UL = umlaut.

differences of ən and *n#*. For the schema cue umlaut+-*e*, we summed up the mean activation differences of all bigraphs and biphones containing an umlaut in word-initial or word-medial position, and the activation differences for the bigraph *e#* or the biphone ə#.

We did not compute the activation differences for monosyllabic stems as it is unclear how this would relate to the other schema cues that are based on only bigrams or trigrams. After calculating the activation differences for the four suffixal (plus umlaut) schema cues, we pitted them against the continuum proposed by Köpcke et al. (2021) (see again Figure 1).

Figure 7 displays the activation differences for the schema cues on a scale that increases from left to right. The top scale shows the activation differences obtained from the bigraph-based model, while the bottom scale presents those obtained from the biphone-based model.

We can observe substantial differences between the NDL-based models: in the bigraph-based models, the plural activation differences of the schema cues are distributed in a range from −0.85 to 0.65, whereas in the biphone-based model, all pertinent activation differences fall below zero. Similarly, in the activation difference scales we find only negative values in the biphone-based model while in the bigraph model we also find positive values.

This means that the written representations can much better discriminate between singular and plural, while the phonological cues all tend towards singular interpretations (though to varying degrees). Interestingly, final <s> or /s/ have very similar activation differences (with −0.85 for the bigraph, and −0.56 for the biphone). All other cues swap position with final <s> or /s/. The reason for this discrepancy between the behavior of <s> or /s/ as against that of the other cues is not clear.

The individual schema cues are distributed with varying distances to each other along the scales for both the bigraph-based and the biphone-based model. For example, *-e* and umlaut + *-e* are located very close to each other, whereas *-en* is positioned more than twice as far apart. Taking a closer look at the schema cues *-e*, umlaut + *-e*, and *-en*, we can see that the relative cue strength for these schema cues broadly reflect the assumptions of the schema cues in both NDL-based models: *-en* emerges as one of the strongest cues for plural, followed by umlaut + *-e*, and then *-e*.

Overall, the results indicate that there is a rather close connection between schema cues and their corresponding bigram-based NDL associations. However, there are remarkable discrepancies, too. These discrepancies have their origin in the fact that the NDL models compute the associations of individual bigrams in the presence of the other bigrams in a given word. In other words, while the NDL models take always the whole word into account, the schema account is exclusively focused on word endings (and umlaut, if any). Hence the possibility of overlap between NDL activations and schema cue strength is limited to begin with.

Interestingly, however, the assumed strength of four of the five schema cues covaries with the word-based NDL associations.

5 Discussion and conclusion

In this paper, we have investigated the question whether the activations of a discriminative learning model are able to predict the number decisions of participants in an experiment on German pseudo-nouns. It turned out that NDL-based models based on bigrams were able to successfully predict participants' decisions. Regression models that used NDL activations as predictors had a slightly better fit than regression models using schema-based cues. These results are in line with the findings on real words by Plag and colleagues (Plag et al., 2024). They found that their NDL models for existing nouns in German were even slightly better in distinguishing plural from singular forms than the schema based predictors.

The present results have important implications for theories of morphological structure. Bigram-based discriminative models do not have explicit morphological information in the input but can still model the morphological effects that are directly encoded in models that use morphemes or suffixes as representations. This means that morphology may be an emergent system in which effects of phonological or morphological units are only by-products of gradient associations between form and meaning.

The NDL models were not without flaw, however. The predictions on word-forms with final /s/ were less accurate than those for other word-forms. Participants had a strong tendency to rate /s/-final words as plural, while the NDL models produced a much lower rate of plural predictions for this set of words. Incidentally, the close connection between final /s/ and plural is also unexpected under the schema account.

Final /s/ has traditionally been found to be controversial in its analysis. The plural suffix *-s* is generally quite rare in the language, but is productively used in certain lexical domains (e.g. with disyllables ending in full vowels and new loanwords). Some researchers advocated its default status (Marcus et al. 1995; Niedeggen-Bartke 1999; Kilbury 2001), others incorporated the *-s* suffix in their systems of smaller or larger generalizations (e.g. Wurzel 1990; Wegener 1992). In general, final /s/ has remained difficult to incorporate in models of German inflection (see, e.g. Marcus et al. 1995 for discussion)

From a dicriminative learning perspective, the low type and token frequency of words that have plural *-s* in the training data may have resulted in rather low plural activations for words in final /s/, which in turn has led to a smaller number of plural predictions in the regression model. It is unclear, however, why the participants showed this plural preference for plural interpretations for /s/-final words.

NDL-based models also allowed us to use bigrams and trigrams to quantify the relative strength of schema cues from a discriminative perspective. The results show that different cue strengths may emerge from a discriminative learning network. The individual NDL-based strengths are, however, not all consistent with the scale of strengths as proposed by the schema account. The discrepancy may in large parts be due to the way the activations are computed in the NDL model (i.e. in the presence of the other bigrams of a given word-form).

The fact that, to a large extent, NDL activations correlate with the purported effects of schema cues may suggest that the effects observed from the perspective of the schema account can be conceived as (at least partly) emerging from discriminative learning.

At a more general level, discriminative learning with its – now demonstrated – capacity to deal with both real words and pseudo-words is a promising alternative to traditional approaches like the schema model. As a first probe into pseudo-words, the present study was restricted to formal cues. As mentioned in the introduction, it is, however, well-known that there are also other determinants at work in the German nominal number system, like gender and semantics. This calls for future work that includes these factors into discriminative learning models that try to come to grips with pseudo-words.

References

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

Arndt-Lappe, Sabine, Robin Schrecklinger & Fabian 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.

Arppe, Antti, Peter Hendrix, Petar Milin, R. Harald Baayen, Tino Sering & Cyrus Shaoul. 2018. *ndl: Naive Discriminative Learning*. https://CRAN.R-project.org/package=ndl. R package version 0.2.18.

Baayen, R. Harald, Yu-Ying Chuang & James P. Blevins. 2018. Inflectional morphology with linear mappings. *The Mental Lexicon* 13(2). 230–268. doi:https://doi.org/10.1075/ml.18010.baa.

Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P Blevins. 2019. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019(1). 1–39. doi:https://doi.org/10.1155/2019/4895891.

Baayen, R. Harald, Petar. Milin, Dusica Filipović Durdević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438–481. doi:https://doi.org/10.1037/a0023851.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1996. CELEX2. *Linguistic Data Consortium, Philadelphia* .

Baayen, R. Harald, Lee H. Wurm & Joanna Aycock. 2007. Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon* 2. 419–463. doi:https://doi.org/10.1075/ml.2.3.06baa.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:https://doi.org/10.18637/jss.v067.i01.

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

Berko-Gleason, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Beser, Deniz. 2021. Falling through the gaps: Neural architectures as models of morphological rule learning. *ArXiv* abs/2105.03710. doi:https://doi.org/10.48550/arXiv.2105.03710.

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.

Brysbaert, Marc. 2022. Word recognition II: Phonological coding in reading. In Margaret J Snowling, Charles Hulme & Kate Nation (eds.), *The Science of Reading: A Handbook*, 79–101. Wiley.

Buch, Achim. 2011. *Linguistic spaces: Kernel-based models of natural language*. Tübingen: Universität Tübingen PhD Dissertation.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Chuang, Yu-Ying & R. Harald Baayen. 2021. Discriminative learning and the lexicon: NDL and LDL. In *Oxford Research Encyclopedia of Linguistics*, Oxford: Oxford University Press. doi:https://doi.org/10.1093/acrefore/9780199384655.013.375.

Chuang, Yu-Ying, Melanie J. Bell, Isabelle Banke & R. Harald Baayen. 2021. Bilingual and multilingual mental lexicon: A modeling study with linear discriminative learning. *Language Learning* 71(S1). 219–292. doi:https://doi.org/10.1111/lang.12435.

Chuang, Yu-Ying, Mihi Kang, Xuefeng Luo & R. Harald Baayen. 2023. Vector space morphology with linear discriminative learning. In Davide Crepaldi (ed.), *Linguistic Morphology in the Mind and Brain*, 167–183. London: Routledge. doi:https://doi.org/10.48550/arXiv.2107.03950.

Chuang, Yu-Ying, Kaidi Lõo, James P. Blevins & R. Harald Baayen. 2020. Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning. In Livia Körtvélyessy & Pavol Štekauer (eds.), *Complex Words: Advances in Morphology*, 119–141. Cambridge: Cambridge University Press.

Daelemans, Walter. 2002. A comparison of analogical modeling of language to memory-based language processing. In Royal Skousen, Deryle Lonsdale & Dilworth B. Parkinson (eds.), *Analogical Modeling: An Exemplar-based Approach to Language*, 157–179. Amsterdam: John Benjamins.

Daelemans, Walter, Jakub Zavrel, Ko van der Sloot & Antal van den Bosch. 2007. *TiMBL: Tilburg Memory Based Learner, version 6.0, Reference Guide: LK Technical Report 04-02.* Tilburg: ILK.

Dankers, Verna, Anna Langedijk, Kate McCurdy, Adina Williams & Dieuwke 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. https://aclanthology.org/2021.conll-1.8.

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

Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Dixon, Maureen, Morag Stuart & Jackie Masterson. 2002. The relationship between phonological awareness and the development of orthographic representations. *Reading and Writing* 15(3). 295–316.

Domahs, Frank, Lisa Bartha-Doering, Ulrike Domahs & Margarete Delazer. 2017. Wie muss ein "guter" deutscher Plural klingen? In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 205–237. Berlin & Boston: De Gruyter Mouton. doi:https://doi.org/10.1515/9783110528978-008.

García-Pérez, Miguel A. 2023. Use and misuse of corrections for multiple testing. *Methods in Psychology* 8. 100120. doi:https://doi.org/10.1016/j.metip.2023.100120.

Grandon, Bénédicte, Marcel Schlechtweg & Esther Ruigendijk. 2024. Processing of noun plural marking in German-speaking children: An eye-tracking study. *Journal of Child Language* 51(6). 1450–1477. doi:https://doi.org/10.1017/S0305000923000521.

Hahn, Ulrike & Ramin 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, Maria, Yu-Ying Chuang & R Harald Baayen. 2021. Modeling morphology with linear discriminative learning: Considerations and design choices. *Frontiers in Psychology* 4929. doi:https://doi.org/10.3389/fpsyg.2021.720713.

Hilpert, Martin. 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, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363.

Kamin, Leon J. 1969. Predictability, surprise, attention, and conditioning. In Byron A. Campbell & Russell M. Church (eds.), *Punishment and Aversive Behavior*, 276–296. New York: Appleton-Century-Crofts.

Kilbury, James. 2001. German noun inflection revisited. *Journal of Linguistics* 37(2). 339–353.

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

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

Köpcke, Klaus-Michael. 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, Klaus-Michael, Sarah Schimke & Verena Wecker. 2021. Processing of German noun plurals: Evidence for first- and second-order schemata. *Word Structure* 14(1). 1–24. doi:https://doi.org/10.3366/word.2021.0173.

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

Kuhn, Max. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5). 1–26. doi:https://doi.org/10.18637/jss.v028.i05.

Marcus, Gary F., Ursula Brinkmann, Harald Clahsen, Richard Wiese & Steven 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, Kate, Sharon Goldwater & Adam Lopez. 2020a. Inflecting when there's no majority: Limitations of encoder-decoder neural networks as cognitive models for German plurals. *ArXiv* abs/2005.08826. doi:https://doi.org/10.48550/arXiv.2005.08826.

McCurdy, Kate, Adam Lopez & Sharon 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. https://aclanthology.org/2020.cmcl-1.8.

Moscoso del Prado Martín, Fermín, R. Bertram, T. Häikiö, R. Schreuder & R. Harald Baayen. 2004. Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30. 1271–1278.

Niedeggen-Bartke, Susanne. 1999. Flexion und Wortbildung im Spracherwerb. In Jörg Meibauer & Monika Rothweiler (eds.), *Das Lexikon im Spracherwerb*, 208–228. Tübingen & Basel: A. Francke.

Nieder, Jessica, Ruben van de Vijver & Fabian Tomaschek. 2023. "All mimsy were the borogoves" – a discriminative learning model of morphological knowledge in pseudo-word inflection. *Langauage, Cognition and Neuroscience* 5(38). 693–710.

Pearce, John M. & Mark 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.

Plag, Ingo, Frank Domahs & Maria Heitmeier. 2024. German nominal number interpretation in an impaired mental lexicon: A naive discriminative learning perspective. *The Mental Lexicon* 18. 417–445. doi:https://doi.org/10.1075/ml.23017.pla.

Plag, Ingo, Frank Domahs & Maria Heitmeier. 2025. Morpho-phonology is not independent of semantics: The case of German nominal number marking. *The Mental Lexicon* doi:https://doi.org/10.1075/ml.24008.pla.

Plag, Ingo, Maria Heitmeier & Frank Domahs. 2022. Modeling German plural recognition in aphasia with discriminative learning. In *20th International Morphology Meeting, September 1-4, 2022, Budapest; 12th Mental Lexicon Conference, October 11-14, 2022, Niagara-on-the-Lake*, https://osf.io/gyfn3/?view_only=af37bc0866394813967cc70cfe09f202.

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

Posit team. 2025. *Rstudio: Integrated development environment for R*. Posit Software, PBC Boston, MA. http://www.posit.co/.

R Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Ramscar, Michael & Daniel 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.

Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957. doi:https://doi.org/10.1111/j.1551-6709.2009.01092.x.

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

Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory*, 64–99. New York: Appleton-Century-Crofts.

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

Saito, Motoki, Fabian Tomaschek & R. Harald Baayen. 2020. Relative functional load determines co-articulatory movements of the tongue tip. In *Proceedings of the 12th in-*

ternational Seminar on Speech Production, New Haven, CT. https://www.researchgate.net/publication/353071272. 217.

Schmitz, Dominic, Ingo Plag, Dinah Baer-Henney & S.D. Stein. 2021. Durational differences of word-final /s/ emerge from the lexicon: Modeling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology* 12. doi:https://doi.org/10.3389/fpsyg.2021.680889. 680889.

Signorell, Andri. 2024. *Desctools: Tools for descriptive statistics*. https://CRAN.R-project.org/package=DescTools. R package version 0.99.58.

Slobin, Dan I. 1973. Cognitive prerequisites for the development of grammar. In Charles Ferguson & Dan I. Slobin (eds.), *Studies of Child Language Development*, 175–276. New York: Holt, Rinehart Winston.

Stein, Simon. D. & Ingo Plag. 2021. Morpho-phonetic effects in speech production: Modeling the acoustic duration of English derived words with linear discriminative learning. *Frontiers in Psychology* 12. doi:https://doi.org/10.3389/fpsyg.2021.678712.

Taft, Marcus & Gail Hambly. 1985. The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language* 24(3). 320–335.

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

Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus & R. Harald 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.

Van de Vijver, Ruben, Emmanuel Uwambayinema & Yu-Ying Chuang. 2024. Comprehension and production of Kinyarwanda verbs in the discriminative lexicon. *Linguistics* 62(1). 79–119.

Wegener, Heide. 1992. Pluralregeln und mentale Grammatik. In Ilse Zimmermann & Anatoli Strigin (eds.), *Fügungspotenzen: Zum 60. Geburtstag von Manfred Bierwisch*, 225–249. Akademie Verlag.

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

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

Wulf, Douglas J. 2002. Applying analogical modeling to the German plural. In Royal Skousen, Deryle Lonsdale & Dilworth 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.

Wurzel, Wolfgang U. 1990. The mechanism of inflection: Lexicon representations, rules and irregularities. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary Morphology*, New York: Mouton de Gruyter.