Morpho-phonology is not independent of semantics: The case of German nominal number marking

Ingo Plag, Frank Domahs & Maria Heitmeier

May 3, 2024

Abstract

Morpho-phonological alternations in inflectional paradigms are commonly analyzed as purely formal phenomena, in which the mapping of phonological structure and morpho-syntactic categories is organized without recourse to semantic properties of the words involved. Some studies have argued, however, that semantics does play a role in inflection, too Baayen & Moscoso del Prado Martín (2005); Shafaei-Bajestan et al. (2022); Chuang et al. (2023). The present paper explores the role of semantics using the Discriminative Lexicon approach (Baayen et al., 2019b).

The test case explored in this paper is German nominal number, a system involving complex morpho-phonological variation, which took center stage in the debates about the architecture of the lexicon, morphology and grammar in the 1990s and early 2000s, and which is still under debate (e.g. Köpcke et al., 2021; Heitmeier et al., 2021; Plag et al., 2024; McCurdy, 2024).

Using word2vec vectors as semantic representations, and triphones as form representations, we created two-layer discriminative learning networks that map form representations directly onto semantic representations (modeling comprehension), and semantic representations onto form representations (modeling production). We first used the LDL mappings between form and meaning to successfully predict the forms and the meanings of the singular and plural nouns taken from a pertinent study (Domahs et al., 2017). Second, a number of measures derived from the network were used to predict the number of a given word-form. The measures very successfully distinguished between singular and plural forms. Plural predictions are favored for word forms that live in a denser semantic network, and whose articulatory trajectory has a stronger association with the word's semantic vector.

Our results demonstrate that semantics, in addition to formal and grammatical properties, may play a decisive role in the representation and processing of German nominal number. The results support the idea that the system of German nominal number can be understood as emerging from the distributional properties of words on the one hand, and basic principles of discriminative human learning on the other.

1 Introduction

Morpho-phonological alternations in inflectional paradigms are commonly analyzed as purely formal phenomena, in which the mapping of phonological structure and morpho-syntactic categories is organized without recourse to semantic properties of the words involved. This standard view of morpho-phonology has been challenged by studies that have shown that the variation in the form of exponents of morpho-syntactic categories is not independent of semantics.

Baayen & Moscoso del Prado Martín (2005) demonstrated that irregular verbs in English, German and Dutch tend to live in denser semantic neighborhoods than regular verbs and have relatively many other semantic neighbors that are morphologically irregular. Shafaei-Bajestan et al. (2022) investigated distributional semantic vectors of English plural nouns and found that these nouns exhibit semantic clusters. For instance, the plurals of words denoting fruit are more similar to each other and less similar to the plurals of other semantic classes. In their study of Russian defective paradigms, Chuang et al. (2023) found that paradigms of inflectionally defective lexemes are characterized by the fact that inflected forms of the same lexeme are semantically less similar to each other, and their meanings are also more idiosyncratic than it is the case for non-defective lexemes.

Such studies raise the questions of how exactly semantics can be included in a model of inflection, and whether morpho-phonological processes (such as inflectional allomorphy), which seem to be by definition purely form-based, are also informed by semantics. In this paper we will address both questions from the perspective of the theory of the 'Discriminative Lexicon' (Baayen et al. 2019b; Chuang & Baayen 2021). This theory takes seriously the idea that morphology is about the mapping of form onto meaning (in comprehension), and meaning onto form (in production), and implements a computational architecture in which relations between forms on the one hand, and meanings on the other, are established by discriminating between different forms and meanings (instead of building them from compositional units, as in morpheme-based morphology, for example). Forms are represented by numerical vectors, and meanings are also represented by numerical vectors. Form and meaning are mapped onto each other by using linear regression. The name of this computational model reflects its two major ideas: 'linear discriminative learning' (LDL).

The test case explored in this paper is German nominal number, a system involving complex morpho-phonological variation, which took center stage in the debates about the architecture of the lexicon, morphology and grammar in the 1990s and early 2000s (e.g. Marcus et al. 1995; Hahn & Nakisa 2000; Penke & Krause 2002). Beyond the inflection on the determiner or the adjective, number in German may be marked on the noun's stem vowel (umlaut) and/or at its end, with one of the suffixes *-[e]n, -er, -e, -s*, or no suffix at all. 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, or no role at all (see Eisenberg & Fuhrhop 2020, pp.171–176, for an overview).

In this paper we use a data set of more than 800 nouns from an independent study (Domahs et al., 2017) to test whether an LDL model can successfully predict the forms of singular and plural nouns from the semantics and vice versa, and whether measures derived from an LDL network are predictive in differentiating singular word-forms from plural word-forms. For the latter task we compare the performance of the LDL model with that of baseline statistical models that predict number on the basis of purely phonological information.

We will demonstrate that semantics, as implemented in a morphological system that maps form and meaning onto each other, does play a decisive role in the nominal number system of German. The results provide support for the idea that inflectional systems are best understood as emerging from the distributional properties of words on the one hand, and basic principles of human learning (i.e. discriminative learning) on the other.

The paper is structured as follows. In the next section we present in more detail previous attempts to model nominal number in German and introduce the reader to discriminative learning models. Section 3 describes the data set and the implementation of the discriminative learning networks. Section 4 contains the results, whose implications are discussed in Section 5.

2 Modeling German nominal number

2.1 Previous approaches

Accounts of the German number system are manifold. Recently, Trommer (2021) has provided a comprehensive subsegmental analysis in which he, however, restricts his analysis to "categorical strictly phonologically motivated generalizations" (p. 616). However, previous research strongly suggests that the formal cues to grammatical number in German operate probabilistically, meaning that virtually no single cue is unambiguous or works in a categorical, deterministic way (Köpcke 1988, 1993; Köpcke et al. 2021).

According to the usage-based schema account of Köpcke and colleagues (1988; 1993; 2021) individual formal cues may combine to constitute a 'schema'. 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 of 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.

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 related 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.

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 across the whole paradigm of case. 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). 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., $Rechen_{\mathrm{Sg/Pl}}$, $Messer_{\mathrm{Sg/Pl}}$). 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 schwa syllables like *-el* (e.g., $Pinsel_{\mathrm{Sg/Pl}}$). 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). 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 more easily perceived than non-syllabic cues – a factor which Köpcke (1998, 1993) called 'iconicity'. As this notion of iconicity can be subsumed under the notion of salience, we will stick with the latter term hereafter. Higher salience of a cue has been shown to improve its learnability and processing in typical (Polišenská 2010; Simoens et al. 2017) and atypical populations (Penke et al. 2016). Yet, the relative importance of these distributional and perceptual factors, their relative impact on processing and their weights in interaction still remain unclear.

To address these concerns, quite a number of studies have applied various

[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).

statistical and machine learning methods to the phenomenon of grammatical number in German nouns in order to model potential human cognitive underpinnings (e.g. Taatgen 2001; Anderson & Lebiere 1998; Hahn & Nakisa 2000; Daelemans 2002; Wulf 2002; Daelemans et al. 2007; Rosen 2022; Buch 2011). More recently, deep learning models have entered the scene (McCurdy et al. 2020a,b; Beser 2021; Dankers et al. 2021), but they tend to be linguistically not very informative, as they aim at prediction accuracy rather than theoretical linguistic understanding. In spite of the fact that the models are quite successful in predicting existing word-forms, the role of the different phonological factors still remains unclear, and semantics is generally not considered.

Most recently, McCurdy (2024) employs a variety of computational models, and compares their performance in experiments to the predictions of a phonologically conditioned baseline. McCurdy concludes that the computational models perform worse than the phonologically-conditioned lexical baseline, which indicates that the models "have not learned the selective feature preferences that inform speaker generalization" (p. iii).

The present study implements a radically different approach that makes crucial use of semantics. We will use a particular implementation of a learning theory, Linear Discriminative Learning (LDL, e.g. Baayen et al. 2018a, 2019b), to explore the role of semantics in inflectional paradigms and to better understand the German number system. In particular, we will explore the idea that semantics may play a crucial, and hitherto underestimated, role in the morphophonology of German nouns.

2.2 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, and numerous studies have shown that it can successfully model many aspects of morphology (see Plag 2018; Lieber 2021 for linguistic textbook introductions). 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 associations (and the resulting representations) are constantly updated based on new, informative experiences. The events associated with each other are called 'cues' and 'outcomes', and the association between cues and outcomes is 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 and a given outcome co-occur. Conversely, the weight decreases whenever a given cue occurs without that outcome. 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. This type of model is known as 'Naive Discriminative Learning' (NDL).

Plag et al. (2024) used NDL networks to model the processing of German number by a patient with primary progressive aphasia (Domahs et al., 2017), and to model the unimpaired system. The authors compared models that used the traditional structural formal cues (kind of determiner, umlaut, suffix, prosodic pattern) borrowed from the schema approach (Köpcke 1988, 1993; Köpcke et al. 2021) with a model that uses segmental and phonotactic information only (biphones, i.e. combinations of two adjacent segments). For the unimpaired system, a model based on only biphones as cues was better able to predict the number of a given word-form than a model using structural phonological cues. Moreover, it was shown that a biphone-based discriminative model trained on the patient's responses was superior to a structure-based model in approximating the patient's behavior.

Based on the tenets of discriminative learning, Baayen and colleagues have developed a theory of the mental lexicon, called the 'Discriminative Lexicon' (see Baayen et al. 2019b; Chuang & Baayen 2021). This theory implements a computational architecture which grew out of NDL, and is called 'linear discriminative learning' (LDL). LDL generates a system of form-meaning relations by discriminating between different forms and meanings (instead of building them from compositional units, as in morpheme-based morphology, for example). In an LDL approach, forms are represented by numerical vectors, and meanings are also represented by numerical vectors.

Form vectors are based either on strings of segmental representations of various lengths (e.g., biphones or triphones), or on representations of acoustic transitions gleaned directly from the speech signal (Arnold et al., 2017; Shafaei-Bajestan et al., 2021). Meaning is taken to be a dynamic concept, being emergent from how words are used, and is represented by semantic vectors, similar to approaches in distributional semantics (e.g. Harris 1954; Firth 1957, see Boleda 2020 for a recent overview). The idea is that, if both forms and meanings can be expressed numerically, we can mathematically connect the two levels of representation, i.e. map meaning onto form, or form onto meaning.

In this system of learning, the two sets of vectors are combined into matrices – a form matrix and a meaning matrix. The form vectors are mapped onto meaning vectors to model comprehension, and meaning vectors are mapped onto form vectors to model production. The mapping between them at the theoretical end-state of learning is estimated using multivariate multiple linear regression (hence the term 'linear discriminative learning'). The network is simple and interpretable, because, in contrast to deep learning networks, it features just two layers (i.e. the form and meaning matrices), both of which are linguistically transparent.

There are two ways of testing and using these networks. One possibility ('internal validation'), is to have the model generate word forms or meanings that can then be compared to empirically observed word forms or meanings. The other possibility of using the networks ('external validation') is to derive secondary measures from the associations given by the networks (Heitmeier, 2022), and to use these measures to predict things outside the network, e.g. independent properties of words, or human behavior. The measures taken from the networks represent different aspects of the association between form and meaning. For instance, one can gauge from the network how much support a particular sequence of sounds receives from the semantics, or how strongly a particular sound sequence activates certain semantic dimensions. Such network measures have been successfully employed to predict inflectional and derivational forms (Baayen et al. 2018a, 2019b; 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 and complex words, including nonce words (Tomaschek et al. 2021; Stein & Plag 2021; Schmitz et al. 2021), the articulatory movements across morphological boundaries (Saito et al. 2020), morphological disorders (Heitmeier & Baayen 2020) and participants' reaction times and decisions in experiments investigating morphological processing (Baayen et al. 2011; Plag et al. 2022; Nieder et al. 2022).

One recent study has applied LDL to German nouns. Heitmeier et al.'s (2021) model successfully produced and comprehended existing case- and number-inflected word-forms on the basis of phonological and semantic information, and the performance on pseudowords seemed to mirror closely that of native speakers. That study, however, did not investigate the role of phonological cues in number inflection, which, according to the morphological literature, is crucial for a theoretical understanding of German nominal number. In the present study we will devise LDL networks to investigate the role of phonological and semantic information in the German number system.

3 Methodology

3.1 The data

The data for the present study were taken from Domahs et al. 2017. Their study aimed at demonstrating the influence of formal cues on the interpretation of German noun phrases for grammatical number in a patient with an impaired mental lexicon due to primary progressive aphasia. 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).

This data set has been used by Plag et al. (2024) for their NDL study. That study, however, only looked at structural-linguistic properties (i.e. phonological cues and the role of determiners), ignoring semantics. The present study crucially includes semantics, to test the potentially additional effect of semantics on the differentiation between singular and plural forms.

In the present study, we do not analyze the patient's responses but the stimuli themselves. The stimuli were specifically selected by Domahs et al. to test the influence of phonological variables. This makes this set of nouns ideal for the present study, in which we explore how semantic properties may influence number allomorphy in German nouns. Given that structural-phonological effects have been established for this data set, we can compare the present LDL account with an account that rests solely on phonological structure.

The original data set contained 944 nominal word-forms, with 472 singular forms and 472 plural forms. Due to the limitations in the availability of semantic vectors for some forms (see below for details), the data set decreased to 830 word-forms, of which 417 are plural forms and 413 are singular word forms. These word-forms represent 428 different lemmas, of which 394 are listed with both a singular and a plural form. We used this data set for all analyses to ensure comparability across analyses. The data set and the scripts of all analyses are available at `https://osf.io/qt8ua/?view_only=522759e9e7a042b58a31c67466221a2e`.

3.2 The baseline models

We established two baseline models that used only phonological variables to predict the number of a given word-form: one model using random forests, the other using generalized linear regression. The phonological variables listed in Table 1 entered our baseline models.

Table 1: Phonological variables used in Domahs et al. (2017).

Variable	**Description**	**Values**
STRESS	stress pattern of the word	`penult/other`
UMLAUT	presence of umlaut	`yes/no`
NUMSYLL	number of syllables	`1, 2`
ENDPOSSPLURAL	word ending is a possible plural suffix	`yes, no`

The phonological variables and the dependent variable are distributed as shown in Table 2.

Table 2: Distribution of phonological variables.

STRESS	`penult`	`other`
	631	199
UMLAUT	`no`	`yes`
	656	174
NUMSYLL	`1`	`2`
	180	650
ENDPOSSPLURAL	`no`	`yes`
	151	679
NUMBER	`plural`	`singular`
	417	413

We ran the random forest analysis in R using the `partykit` package and its `cforest()` function with default settings (Hothorn & Zeileis 2015), with the four predictors from Table 1 and number as dependent variable. The regression model was fitted using the `glm` function, and the `step` function for model simplification. The two predictors STRESS and NUMSYLL were highly correlated (*rho=0.94*). To avoid collinearity issues, we fitted two different models, each with only one of the two correlating variables. We report below the model that included NUMSYLL instead of STRESS, as this model had a lower AIC value and was significantly better in a log-likelihood test.

3.3 The LDL model

The general workflow involved the following steps:

1. Train an LDL model on German nouns.
2. Use the matrices of the model to predict word forms and to predict meanings.
3. Extract production and comprehension measures from the LDL network.
4. Use these measures to predict number for the word-forms in the data set.

To train the LDL network we needed two matrices, one of them representing the meaning of the words in the lexicon (including our target words), and the other one representing their form. Figures 1 and 2 give a toy example each of a form matrix and a semantic matrix, respectively (from Heitmeier et al. 2021, p.5). In an LDL implementation these matrices are mapped onto each other using the matrix algebra of multivariate multiple regression (see, for example, Chuang & Baayen (2021), in which the mathematical underpinnings of LDL implementations are described in detail). In the present paper, triphones are used as cues (illustrated as `cue1` through `cue5` in Figure 1), and word2vec embeddings as semantic vectors (illustrated with the dimensions `S1` through `S5` in Figure 2).

$$\mathbf{C} = \begin{array}{c} \\ \text{wordform1} \\ \text{wordform2} \end{array} \begin{array}{c} \begin{array}{ccccc} \texttt{cue1} & \texttt{cue2} & \texttt{cue3} & \texttt{cue4} & \texttt{cue5} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Figure 1: Toy **C** matrix with phonological cues.

$$\mathbf{S} = \begin{array}{c} \\ \text{wordform1} \\ \text{wordform2} \end{array} \begin{array}{c} \begin{array}{ccccc} \texttt{S1} & \texttt{S2} & \texttt{S3} & \texttt{S4} & \texttt{S5} \end{array} \\ \begin{pmatrix} 0.1 & 0.004 & -1.95 & 0.03 & -0.54 \\ -0.49 & -0.32 & 0.03 & 1.06 & 0.98 \end{pmatrix} \end{array}$$

Figure 2: Toy **S** matrix with semantic vectors

For the meaning matrix we wanted to use the distributional vectors from Heitmeier et al. (2021). Since there was a problem of overlap of their data set and Domahs et al.'s which would have led to a rather small data set to eventually analyze, we enlarged the list of word-forms in Heitmeier et al.'s data set by including more forms from CELEX: We used all monomorphemic nouns (according to CELEX), plus the full paradigms of the nouns in the Domahs et al. stimuli list. For the word-forms in this list we employed the semantic representations that were also used in Heitmeier et al. 2021, i.e. 300-dimensional vectors generated with word2vec, trained on the German Wikipedia (Yamada et al. 2020). This resulted in a list of 30,859 word-forms of 4,032 unique lemmas (amounting to an average of 7.65 word-forms per lemma, which is close to the theoretical 8 paradigm cells per noun lemma in German). The word-forms contain 10,186 unique phonological forms (10,323 unique orthographic forms), i.e. each unique word-form occurs about three times in the dataset. This reflects the abundant homophony in the German noun system. Semantic vectors were still not available for all word forms in the Domahs et al. data set, which resulted in the loss of 12.1 percent of the Domahs et al. data. The final data set had 830 word forms. For representing words' forms, we used triphones, which resulted in a matrix with 30,859 rows (word forms) and 5,950 columns (cues).

After fitting of the network we first used the transformation matrices to predict word-forms from the semantic vectors, and to predict semantic vectors from the word-forms. We compared the predicted meanings with the observed meanings, and the predicted forms with the real forms, to see how well the predicted LDL-based mappings between form and meaning are able to approximate the language system. The results of these analyses are described in section 4.1.

For predicting word-forms an additional algorithm is needed that concatenates the triphones in the right order, to estimate the word's articulatory path. This estimation is done in the following way: first, trigrams with low support from the semantics below a certain threshold are simply discarded. Second, all possible remaining paths are calculated, and, third, that path is selected for which the corresponding predicted semantic vector best matches the semantic vector that is the target for articulation (see Baayen et al. 2018a, 2019b; Chuang et al. 2020, for details and discussion.).

We then extracted a range of production and comprehension measures from the network in order to investigate how predictive these measures are for determining a noun's number. The measures are listed in Table 3 and explained and discussed in detail below. We chose measures that have been successfully used in previous work to investigate morphological problems (see again the references given above). The comprehension measures are semantic in nature and quantify semantic properties of the target word and its relation to other words as predicted from the form matrix. The production measures are phonetic/phonological in nature and quantify properties of the target's form and its relation to other forms as predicted from the semantic vectors.[2]

Table 3: LDL measures

LDL measure	**Description**
Semantics/comprehension	
L1Norm	City block distance of $\hat{\mathbf{S}}$
L2Norm	Euclidean distance of $\hat{\mathbf{S}}$
Density	Semantic Density (correlation-based)
EDNN	Euclidean Distance to Nearest Neighbour
NNC	Nearest Neighbor Correlation
ALC	Average Lexical Correlation
Phonetics-phonology/production	
Support	Support in the form matrix for the last triphone of each word
PathSum	Summed support for the predicted path at each timestep
MeanWordSupport	Summed path support divided by each word form's length
PathCounts	Number of paths detected (for the thinned graph)
ALDC	Average Levenshtein Distance of all candidate productions

L1Norm and L2Norm: Both measures compute the length of the predicted semantic vector $\hat{\mathbf{S}}$ of a target form. The L1Norm is the sum of the absolute values of vector elements of a given word's predicted semantic vector, i.e., its city-block distance. In contrast, the L2Norm is computed as the square root of the sum of the squared values of $\hat{\mathbf{S}}$, i.e., its Euclidean distance. Higher values of L1Norm and L2Norm indicate stronger links to many other words' meanings.

[2]An overview over all available measures can be found at `https://mariahei.github.io/JudiLingMeasures.jl/dev/#Overview-over-all-available-measures`.

Density: This measure is derived by computing the mean correlation of the target's $\hat{\mathbf{S}}$ with the semantic vectors of its top eight neighbors in $\mathbf{S}$ in terms of Pearson correlation. The higher Density, the more similar these words are, indicating a denser semantic neighbourhood.

ALC: Being another measure of semantic relatedness, Average Lexical Correlation is the mean value of all correlation values of a target's estimated semantic vector $\hat{\mathbf{S}}$ with all other semantic vectors in $\mathbf{S}$. Higher ALC values indicate a stronger connection of the target word's meaning to the words' meanings in the lexicon at large.

EDNN: This variable gauges the Euclidean Distance between a target's $\hat{\mathbf{S}}$ and its nearest neighbour in $\mathbf{S}$ in terms of Euclidean distance. A lower value (i.e. smaller distance) indicates that this target's nearest neighbor is semantically very similar. EDNN thus measures another aspect of semantic neighbourhood.

NNC: The Nearest Neighbour Correlation is an alternative way of measuring the semantic distance of a target and its closest semantic neighbor. Instead of the Euclidean distance, one uses the highest correlation value of the target's $\hat{\mathbf{S}}$ with any of the other semantic vectors in $\mathbf{S}$. The highest value identifies the nearest neighbor and estimates how close the target is to its nearest neighbor. Like EDNN, NNC can be interpreted as a measure of semantic similarity between a target and its nearest neighbor.

Support: This measure describes the amount of semantic support the word-final trigram, (i.e., `ks#`, `en#`) obtains for each target word. Thus, Support is simply the activation the trigram receives in the predicted form vector $\hat{\mathbf{c}}$. Higher values of this variable indicate a higher semantic support for the word-final triphone which potentially includes the segments of interest, i.e. those that are strongly associated with plural, i.e. exponents of suffixes.

PathSum: This measure is based on the probabilities of the transitions from one triphone to the next, i.e. the support of these transitions from the network. These transitions are referred to as an articulatory 'path', which is computed by the production model. PathSum is the summed support that the triphones in the predicted articulatory path receive at each timestep from the semantics. The measure can be interpreted as a measure of phonotactic or articulatory certainty, with higher values indicating a higher certainty.

MeanWordSupport: This measure is based on PathSum but puts it in relation to a word's length. This makes sense because longer words have more transitions, which means that PathSum increases not only with higher support, but also with increasing number of transitions, i.e. increasing length of words. By dividing the value of PathSum by the number of transitions we control for path length and arrive at the average transition support in each word. The resulting measure can be seen as an alternative measure of articulatory certainty. The higher the average transition probabilities in a given target word, the more certain the speaker will be in pronouncing this word, based on its semantics.

PathCounts: This variable gives the number of paths, i.e., possible sequences of triphones, that the production model has detected based on the activations in the network. Higher values indicate the existence of multiple paths, i.e. candidate forms for production. Hence, PathCounts can be interpreted as a measure of phonotactic uncertainty.

ALDC: The Average Levenshtein Distance of all candidate forms is derived by computing the mean of all Levenshtein distances of a target word form and its competing candidate forms. If a target word has only one candidate form, the Levenshtein distance between the two forms is its ALDC. If there are more candidate forms, the mean of the individual Levenshtein distances between candidates and target form constitutes the ALDC. This measure indicates the similarity between a target's candidate forms and the target and can thus be taken as a measure of phonological neighbourhood density. A large value means that the target lives in a sparse phonological neighbourhood.

The LDL measures just described were added to the Domahs et al. (2017) data set. We then applied different analyses to relate these measures to the grammatical number of the nouns. In order to see whether the LDL measures can differentiate between singular forms and plural forms, we employed t-distributed Stochastic Neighbor Embedding (t-SNE, Van der Maaten & Hinton 2008), which is a nonlinear dimensionality reduction technique that reduces the high-dimensional space of our LDL measurements to two dimensions that can then be straightforwardly visualized. To statistically corroborate the t-SNE results we employed linear discriminant analysis (LDA) (see, for example, Shafaei-Bajestan et al. 2022; Schäfer 2023). LDA allowed us to quantify and assess the quality of the model's predictions. Principal component analyses were used for further exploration of the properties of individual LDL measures. The results of these analyses are presented in section 4.2.2

4 Results

In this section we will describe the results of the different analyses. We will first look at the prediction of word forms and meanings before turning to the analyses that were used to predict the number of a given word-form, starting with the baseline models and finishing with the models that used LDL-based measures.

4.1 Predicting form and meaning using LDL

Let us first investigate comprehension, i.e. how well the semantic vectors predicted by the model approximate the target vectors. For this task, one can look at the correlation of the predicted vector with observed vectors. Ideally, the observed vector is the vector that has the highest correlation with the predicted vector. Applying this very strict evaluation metric, and comparing the predicted vector with the vectors of all 30,859 words in the training set, we get an accuracy of 78 percent. If we relax our criterion and check whether the target word's vector is among the 5, 10 or 20 nearest neighbors, the accuracies rise to 88, 95 and 97 percent, respectively. These figures are comparable to those of previous pertinent studies (e.g. Baayen et al., 2018b, 2019a; Heitmeier et al., 2021) and indicate that the model is quite successful in relating form to meaning in German nouns.

In production, things are quite similar. Comparing the predicted word-form with all 30,859 word-forms in the training set, we get an accuracy of 85 percent when looking for the same form. In sum the LDL transformation matrices are quite successful in mapping forms onto meanings and meanings onto forms.

Focusing on the subset of our data which is in the data set by Domahs et al. (2017), we find that the model correctly comprehends 86.6% of all word-forms (singulars: 82.6%, plurals: 90.6%) and produces 92.5% of all word-forms correctly (singulars: 92.7%, plurals: 92.3%).

We will now zoom in on number marking, testing whether measures gleaned from the LDL matrices are able to discriminate between plural and singular forms. We will first look at baselines models that are based on structural phonological features that are traditionally held to differentiate singular and plural forms.

4.2 Predicting number

4.2.1 The baseline models: Predicting number based on structural-phonological properties

The random forest model reached an accuracy of 69 percent (F1 score: 0.75[3]). It had a major problem in classifying singular word-forms correctly. Of the 413 singulars in the data, 229 were misclassified as plurals (55.4%). Of the 417 plural forms, in contrast, the model predicted 391 correctly as plurals (93.8%). This means that plurals can indeed be characterized by phonological characteristics, but that these characteristics are also shared by many singular forms.

We obtained exactly the same classification is obtained when using the generalized linear model with UMLAUT, NUMSYLL and ENDPOSSPLURAL variables. If we turn the logits predicted by the regression model into binary decisions, these decisions are the same as the ones by the random forest model.

4.2.2 Predicting number using the LDL model: t-SNE and LDA analysis

The t-SNE analysis is able to very clearly separate singulars from plurals. This is shown in Figure 3. The plural forms cluster in the upper part of the plot, a bit skewed to the left, while the singulars are clustered in the bottom part, a bit skewed to the right. One small cluster of singulars can be seen in the upper left corner of the graph (tSNE1 $<$ -5, tSNE2 $>$ 15), suggesting that this cluster shares certain properties with the plurals that they do not share with the rest of the singular forms, which are all located on the opposite side of the graph.

[3]The F1 score is a measure of the harmonic mean of precision and recall

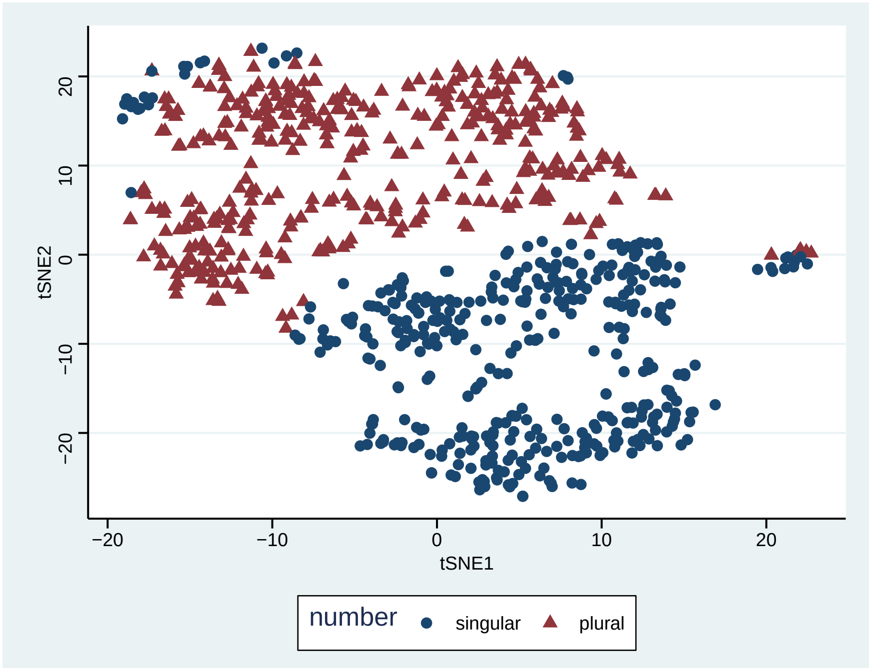

Figure 3: Separation of singulars and plurals based on LDL measures, using t-SNE.

If we inspect this aberrant cluster of singulars more closely, we see that the words are indeed very similar to each other in form. The words are plotted again in Figure 3, which zooms in on the pertinent area of Figure 4.

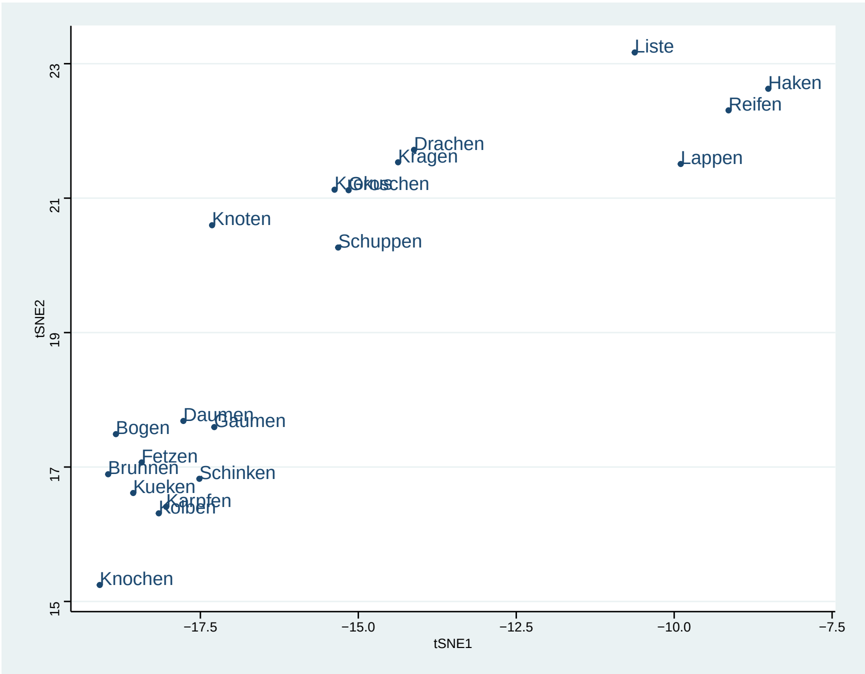

Figure 4: Cluster of aberrant singulars, enlarged from Figure 3.

These 20 singulars (*Bogen, Brunnen, Daumen, Drachen, Fetzen, Gaumen, Groschen, Haken, Karpfen, Knochen, Knoten, Kolben, Kragen, Krokus, Küken, Lappen, Liste, Reifen, Schinken, Schuppen*) share the following formal properties:

- All end in a possible plural suffix (18 times *-en*, one time *-e*, one time *-s*).
- All have two syllables.
- All have penultimate stress.
- All but three (*Krokus, Bogen, Liste*) have the same form as their plurals.
- All but one (*Küken*) have no umlaut.

With the first three properties, the singular forms in the two clusters exhibit phonological properties that are typical of plurals, i.e. formally, they are poor singulars. The clustering according to the five properties is remarkable, because the LDL model has no explicit information about these properties. The clustering of properties can thus be considered as emergent from the learned relationships between form and meaning in the lexicon, with ‘form’ referring to sub-lexical sequences of three segments (or two segments plus word margin).

There is another remarkable cluster of forms, on the far right of Figure 3 (tSNE1 $>$ 19, tSNE2 around zero), set off from the singulars, and comprising both singular and plural forms. A closer inspection reveals a similar pattern as we saw with the other cluster. The singular forms all share a number of phonological properties: they all end in schwa (apart from one ending in *-s*), are bisyllabic trochees, and (apart from one) have an umlaut in their stem. The plural forms also feature two syllables, umlaut (apart from one form), and end in the schwa plural suffixes *-en* or *-e*. The words are plotted in Figure 5, which zooms in on the pertinent area of Figure 3.

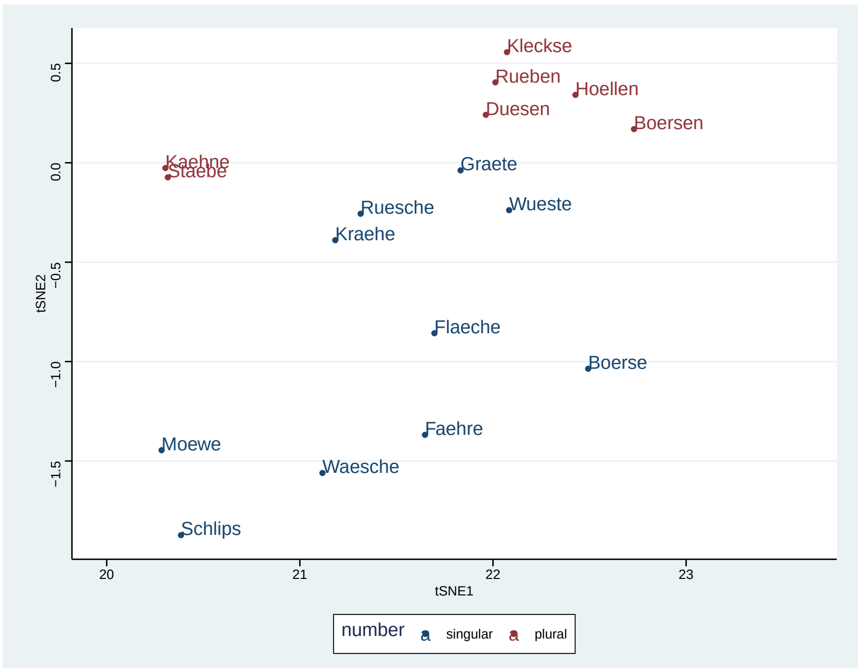

Figure 5: Cluster of peculiar singulars and plurals, enlarged from Figure 3.

To corroborate the t-SNE results we also fitted a linear discriminant analysis. We used the same predictors in the LDA analysis as in the t-SNE analysis presented above. LDA yielded very similar results, with an accuracy of 0.77 (F1 score: 0.79)

4.3 Inspecting individual LDL measures: PCA analysis

In order to inspect more closely the relation of the individual LDL measures with number, regression models are the method of choice. However, the LDL measures bring about issues of collinearity because there are strong correlations among the comprehension measures (Figure 6) and among the production measures (Figure 7). Each comprehension measure correlates strongly with all other comprehension measures. Of the production measures, PATHCOUNTS and

ALDC do not correlate highly with either of the other two production measures (which correlate strongly with each other).

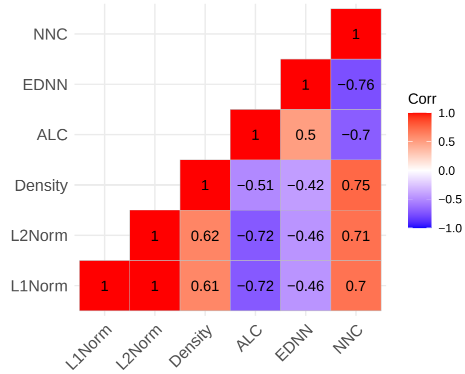

Figure 6: Correlation Matrix for semantic/comprehension measures

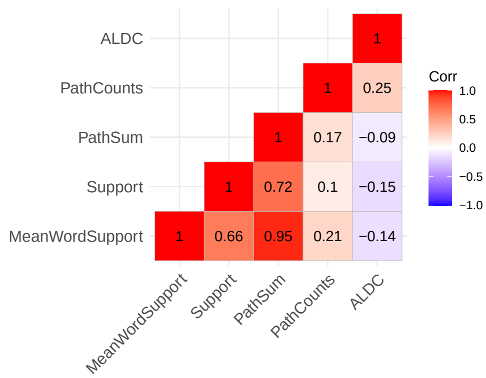

Figure 7: Correlation Matrix for phonological/production measures

To address these issues of potential collinearity we used principal component analysis (see, e.g., Baayen 2008), implementing the `pcrcomp` function and the `factoextra` package (Kassambara & Mundt, 2020) in R. In a principal component analysis, the dimensionality of the data is reduced by transforming the different variables into linear combinations that are orthogonal to each other. These uncorrelated new linear predictors are called ‘principal components’. All predictor variables have been centered before entering the principal component analysis (and we have added a ‘_C’ to their names to indicate this transformation).

The first four principal components explain 0.83 of the variance and are selected here for further inspection and interpretation. The factor loadings are given in Table 4.

Table 4: Factor loadings for the centered LDL predictors in the PCA model (centering is indicated by the '_C' following the variable name).

	PC1	PC2	PC3	PC4
Comprehension				
L1Norm_C	0.42	0.08	-0.11	-0.34
L2Norm_C	0.42	0.08	-0.11	-0.34
Density_C	0.38	-0.08	0.16	0.10
NNC_C	0.42	0.15	0.19	0.27
EDNN_C	-0.32	-0.19	-0.23	-0.60
ALC_C	-0.37	-0.19	-0.03	0.26
Production				
MeanWordSupport_C	0.19	-0.53	-0.06	-0.04
Support_C	0.08	-0.52	-0.01	0.24
PathSum_C	0.18	-0.55	-0.09	0.04
PathCounts_C	0.06	-0.06	-0.41	-0.17
ALDC_C	0.10	0.19	-0.82	0.41

The principal components can be interpreted as follows: PC1 has its highest loadings for the comprehension measures and thus clearly reflects semantic interpretation based on the wordform. The first four measures have positive loadings, the last two negative loadings, which indicates that they tap into different dimensions of the semantic properties. This is to be expected since higher values of L1Norm_C, L2Norm_C, Density_C and NNC_C all indicate a higher degree of semantic co-activation. EDNN_C also indicates semantic co-activation but has the opposite sign (i.e. a larger Euclidean distance indicates a smaller amount of co-activation). ALC_C also has the opposite sign, which indicates that a denser network of close semantic neighbours (as indicated by higher values of, for instance, Density_C or NNC_C), goes together with lower values of ALC_C. This means that words living in a denser semantic neighborhood are on average less semantically similar to all the other words in the lexicon.

PC2 taps into production. All measures of support for the articulation (MeanWordSupport_C, Support_C, PathSum_C) have high negative loadings. PC3 also reflects production. PathCounts_C and ALDC_C, which are both measures of articulatory uncertainty have high negative loadings.

PC4 combines the high influence of a measure of semantic co-activation, i.e. EDNN_C, with the high influence of a production measure, ALDC_C. The negative loading of EDNN_C means that an increase in PC4 indicates more semantic co-activation. At the same time, the positive loading of ALDC_C indicates that a higher value of PC4 goes together with sparser phonological neighborhoods.

In order to see how these measures relate to number, we fitted generalized regression models with the principal components as predictors. Number was used as the dependent variable. We used the `step` function from the `MASS` package in R to simplify models.

Table 5: Final regression model for grammatical number, using the first four principal components. *p<0.1; **p<0.05; ***p<0.01)

	coefficient	significance	standard error
PC1	0.192	***	0.040
PC2	-0.628	***	0.060
PC4	0.693	***	0.100
Constant	0.087	***	0.082
Observations	822		
Log Likelihood	-465.064		
Akaike Inf. Crit.	938.128		

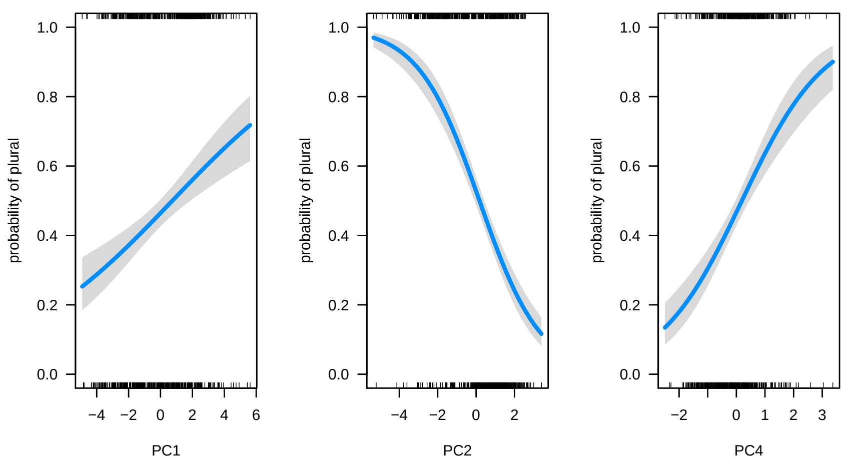

Figure 8: Partial effects of the final regression model, using the first four principal components in the initial model.

The final model still contains PC1, PC2 and PC4. We can see in Table 5 and in Figure 8 that PC2 has the strongest effect. To interpret the effect we need to keep in mind that the predictors underlying PC2 are negatively correlated with this component. The negative slope of the regression line indicates that more semantic support for the word-form goes together with a higher chance of a form being plural. Higher values of PC1 go together with plural, which means that plural word forms live in a denser semantic space than singulars.

In other words, plural words are semantically closer to each other than singular words. Both effects of semantics are not really surprising as plurals share an important meaning component: 'more than one'. Finally, the effect of PC4 can be interpreted in such a way that plurals tend to have closer semantic neighbors and a less dense phonological neighborhood. This makes intuitive sense, as we observed a similar effect for PC1. With regard to the phonological aspect of this component, the effect is also not surprising, as plurals are typically characterized by additional segments (i.e. a suffix), which inevitably increases the chances of a higher average Levenshtein distance to other forms in the lexicon.

The model with PC1, PC2 and PC4 as predictors has a C-value of 0.80. If we turn the predicted probabilities into binary decisions (with a threshold of probabilities higher than 0.5 being recoded as `plural`, all others as `singular`), the accuracy is 0.73, the F1 score equals 0.76. The accuracies for singular forms and plural forms respectively are 0.69 correct for singulars, 0.80 correct for plurals.

If we use all principal components (analogous to the t-SNE and LDA analyses above, where we used all LDL predictors) the model fit improves (accuracy: 0.78, F1 score: 0.79 C: 0.86; 0.75 correct for singulars, 0.82 for plurals). In summary we can say that the LDL comprehension and production measures are indeed highly predictive of the grammatical number of a given word-form.

5 Discussion and conclusion

In this paper we have used a data set from an independent study to investigate whether it is possible to differentiate between singular and plural word forms on the basis of a discriminative learning network. In this network, matrices of form representations (i.e. triphones) and meaning representations (i.e. word embeddings) are mapped onto each other to model production and comprehension. With such a network we tested whether, as it is commonly assumed, morpho-phonological variation in paradigms is purely a matter of form, or may also involve semantics.

We first investigated whether the model is able to predict the correct inflected form of a word from its semantic vector (mimicking production), and the correct semantics from its form (i.e., comprehension). The LDL model was quite successful in predicting forms and meanings, and did so with almost equal success for both singulars and plurals. The model thus is able to comprehend and produce inflected German nouns with surprising accuracy, given that the architecture has no access to morphological information. The rather simple mapping of triphones and semantic vectors can achieve this.

We then demonstrated that the baseline statistical models with exclusively phonological variables as predictors were able to predict the number of a form successfully only for plural forms, while large proportions of singular forms were wrongly classified. In contrast, the measures gleaned from the LDL network are generally more successful in differentiating plural and singular forms. Moreover, these measures perform almost equally well for both plural and singular forms. Furthermore, it was found that the measures cluster in ways that reflect phonological patterns, without these patterns being explicitly encoded in the network.

A point we want to raise here is how our current LDL implementation compares to a similar implementation in NDL done by Plag et al. (2024). At least theoretically, since they employ the same learning mechanism (i.e. discriminative learning), such a comparison should be instructive as to the leverage that semantics may provide over purely form-based cues. However, in reality, the conceptualisations of NDL and LDL are entirely different: while Plag et al. (2024) analysed the current data as a learning problem where form cues are mapped to a singular and plural outcome (i.e. the model is trained to explicitly predict whether a word-form is a singular or a plural), the present LDL implementation aims to model lexical processing of word-forms more generally, i.e. how word-forms' meanings can be predicted given their form and vice versa. The distinction between singulars and plurals arises as an epiphenomenon rather than being explicitly learned. It is therefore of limited interest to compare the performance of the two kinds of model. Instead, the present LDL model should be seen as proof of concept that distinctions between singulars and plurals of German nouns can arise in a model which is not trained to distinguish them.

The findings presented in this paper have three important theoretical implications. First, inflection is not only about categories of form and morphosyntactic specifications. And morpho-phonological alternations constitute not only patterns of variable form, but reflect the mapping of form and meaning. Morpho-phonological alternations are thus phenomena at the very heart of morphology, which is the mapping of form and meaning at the word-level.

Second, morpho-phonological generalizations can emerge in a discriminative computational system without any explicit knowledge of the phonological properties involved in these generalizations. For instance, we found generalizations over stress, the number of syllables, and word-endings (i.e. 'suffixes'), even though the model only worked with information of sequences of three segments. Morphology is gradient and emergent from the relations of form and meaning in the lexicon.

Third, the successful internal and external validations of the LDL model have demonstrated that discriminative learning networks can successfully model important properties of complex words, including morpho-phonological patterns found in inflectional paradigms. The present results thus also lend further support to discriminative approaches to language and language learning.

References

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

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.

Arnold, D., F. Tomaschek, F. Lopez, K. Sering & R. H. Baayen. 2017. Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLOS ONE* 12(4). e0174623. URL `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174623`.

Baayen, R. H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Y.-Y. Chuang & M. Heitmeier. 2019a. *WpmWithLdl: Implementation of Word and Paradigm Morphology with Linear Discriminative Learning.* R package version 1.4.6, available at `https://osf.io/xq92s/`.

Baayen, R. H., Y.-Y. Chuang & J. P. Blevins. 2018a. Inflectional morphology with linear mappings. *The Mental Lexicon* 13(2). 230–268. doi: 10.1075/ml.18010.baa.

Baayen, R. H., Y.-Y. Chuang, E. Shafaei-Bajestan & J. P. Blevins. 2019b. 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: 10.1155/2019/4895891.

Baayen, R. H., Y. Chuang & J. P. Blevins. 2018b. Inflectional morphology with linear mappings. *The Mental Lexicon* 13(2). 232–270.

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. & F. Moscoso del Prado Martín. 2005. Semantic density and past-tense formation in three Germanic languages. *Language* 81. 666–698.

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`.

Boleda, G. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6(1). 213–234. doi: 10.1146/annurev-linguistics-011619-030303.

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. & R. H. 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, Y.-Y., D. Brown, H. Baayen & R. Evans. 2023. Paradigm gaps are associated with weird "distributional semantics" properties. *The Mental Lexicon* doi: 10.1075/ml.22013.chu.

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`.

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

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.

Firth, J. 1957. A synopsis of linguistic theory 1930-1955. *Studies in Linguistic Analysis* 1–32.

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.

Harris, Z. 1954. Distributional structure. *Word* 10(2-3). 146–162. doi: https://doi.org/10.1080/00437956.1954.11659520.

Heitmeier, M. 2022. JudiLingMeasures.jl. `https://github.com/MariaHei/JudiLingMeasures.jl`.

Heitmeier, M. & R. H. Baayen. 2020. Simulating phonological and semantic impairment of English tense inflection with Linear Discriminative Learning. *The Mental Lexicon* 15(3). 385–421. doi: https://doi.org/10.1075/ml.20003.hei.

Heitmeier, M., Y.-Y. Chuang & R. H. 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, 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.

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.

Kassambara, A. & F. Mundt. 2020. Extract and Visualize the Results of Multivariate Data Analyses [R package factoextra version 1.0.7], URL `https://api.semanticscholar.org/CorpusID:225959968`.

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`.

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

Van der Maaten, L. & G. Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9(11).

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. 2024. Rules, frequency, and predictability in morphological generalization: behavioral and computational evidence from the German plural system .

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`.

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.

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.

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

Plag, I., M. Heitmeier & F. 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*, URL `https://osf.io/gyfn3/?view_only=af37bc0866394813967cc70cfe09f202`.

Plag, I., M. Heitmeier & F. Domahs. 2024. German nominal number interpretation in an impaired mental lexicon: A naive discriminative learning perspective. *The Mental Lexicon* .

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.

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`.

Saito, M., F. Tomaschek & R. H. Baayen. 2020. Relative functional load determines co-articulatory movements of the tongue tip. In *Proceedings of the 12th international Seminar on Speech Production*, New Haven, CT. URL `https://www.researchgate.net/publication/353071272`. 217.

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.

Schmitz, D., I. Plag, D. Baer-Henney & S. 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: 10.3389/fpsyg.2021.680889. 680889.

Schäfer, M. 2023. The role of meaning in the rivalry of -ity and -ness: Evidence from distributional semantics. *submitted* .

Shafaei-Bajestan, E., M. Moradipour-Tari, P. Uhrig & R. H. Baayen. 2021. LDL-AURIS: A computational model, grounded in error-driven learning, for the comprehension of single spoken words. *Language, Cognition and Neuroscience* 1–28. doi: 10.1080/23273798.2021.1954207.

Shafaei-Bajestan, E., M. Moradipour-Tari, P. Uhrig & R. H. Baayen. 2022. Semantic properties of English nominal pluralization: Insights from word embeddings.

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.

Stein, S. D. & I. 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: 10.3389/fpsyg.2021.678712. 678712.

Taatgen, N. A. 2001. Extending the past-tense debate: A model of the German plural. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 23, 94–108.

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.

Trommer, J. 2021. The subsegmental structure of German plural allomorphy. *Natural Language & Linguistic Theory* 39(2). 601–656. doi: https://doi.org/10.1007/s11049-020-09479-7.

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.

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.

Yamada, I., A. Asai, J. Sakuma, H. Shindo, H. Takeda, Y. Takefuji & Y. Matsumoto. 2020. Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia. In Q. Liu & D. Schlangen (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 23–30. Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-demos.4. URL `https://aclanthology.org/2020.emnlp-demos.4`.