Learn What is Detectable, Detect What is Useful

Sergei Monakhov (sergei.monakhov@uni-jena.de)
Friedrich-Schiller-Universität Jena,
Fürstengraben 1, 07743 Jena, Germany

Holger Diessel (holger.diessel@uni-jena.de)
Friedrich-Schiller-Universität Jena,
Fürstengraben 1, 07743 Jena, Germany

Brisca Balthes (briscab@lst.uni-saarland.de)
Saarland University,
Campus, 66123 Saarbrücken, Germany

Abstract

Many computational models of morphology represent complex words by n-grams to account for lexical processing and acquisition. However, while n-gram models are simple and efficient, they are not without problems. From a cognitive perspective, it is unclear how n-gram words are represented in the mental lexicon and how these representations affect language use and acquisition. From a computational perspective, these models are problematic because n-gram representations are often ambiguous and redundant: they make very limited use of distributional information and neglect the role of efficiency and sequential processing in language use and acquisition. In this paper, we present a new computational approach to morphology that is cognitively more plausible than standard n-gram models. By analyzing data from the nominal number system in German, we show that a task-specific algorithm of linear processing guided by the principles of efficiency and reliability outperforms state-of-the-art n-gram models and also makes predictions about lexical processing that are consistent with the judgments of German native speakers in a psycholinguistic experiment.

Keywords: lexical models; grammar network; graph theory; shallow processing; morphological structure; German plural

Introduction

The usage-based theory of language posits words, not morphemes, as the basic units of lexical processing (Matthews, 1991; Bybee, 1985; Blevins, 2016). However, given that speakers are able to form novel complex words (and that listeners are able to understand them), it is reasonable to assume that lexical representations include units smaller than the word (e.g., classical morphemes, phonaesthemes, phonological chunks). According to the word-based model, these units are parsed out of individual lexemes based on statistical regularities in the input data (Plaut & Gonnerman, 2000; Rastle, Davis, & New, 2004; Amenta & Crepaldi, 2012). The word-based model of morphology has become very popular in psycholinguistic research on the mental lexicon (Smolka, Preller, & Eulitz, 2014; Milin, Smolka, & Feldman, 2017; Smolka, Libben, & Dressler, 2019) and in the usage-based theory of grammar (Bybee, 1995, 2001, 2007; Booij, 2010, 2016; Diessel, 2019; Jackendoff & Audring, 2020). There is now a large body of psycholinguistic research showing that statistical learning plays a central role in language acquisition and processing (Saffran, 2001; 2002; McCauley & Christiansen, 2019).

Consistent with this research, cognitive scientists have developed various computational models in which morphological structure is derived from statistical patterns in large corpora, e.g.: (i) the generalized context model (Nosofsky, 1990), (ii) the analogical model (Skousen, Lonsdale, & Parkinson, 2002), (iii) artificial neural network models (Rumelhart & McClelland, 1987; Seidenberg, 2005), and (iv) the discriminative learning model (Baayen et al., 2019). Albeit very different in terms of implementation, these models are conceptually similar in that they view learning as a classification problem in which new stimuli are assigned to categories based on reliable cues that are associated with one of the outcomes.

The analysis of language cues is of central significance to statistical grammar learning. But what exactly are language cues? Traditionally, language cues have been defined as standard linguistic features, such as case marking and word order, that help the child to learn language and that adult speakers use to determine the meaning and structure of words and phrases (MacWhinney, 1987; Köpcke, 1988; 1998). Since all language cues can be mapped onto more than one pattern, they have different information values, depending on their properties, such as frequency, detectability, availability, and reliability. All these properties have been shown to influence language use and L1 acquisition (Pescuma et al., 2021; Grandon et al., 2023).

However, since the analysis of language cues requires a great deal of linguistic theorizing and hand-coding, this approach is not easy to implement in computational models. Besides being inefficient, it complicates generalisation to understudied languages and renders the entire procedure of analogical modelling not particularly rigorous. Moreover, it has been shown that traditional linguistic cues do not (necessarily) improve the performance of simple n-gram models (MacWhinney et al., 1989; Plag, Heitmeier, & Domahs, 2023).

Given the simplicity and efficiency of n-grams, it does not come as a surprise that most computational models of morphology rely on these representations. However, while

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

n-gram models are simple and powerful, they have their own challenges. In particular, they (i) appear cognitively implausible, (ii) neglect the cue detectability property, (iii) abandon the idea of constrained statistical learning, and (iv) make (almost) no use of distributional information.

First, n-gram models are difficult to interpret from a cognitive perspective. It remains unclear exactly what, according to these models, is represented in the mental lexicon and how speakers acquire n-grams in the process of language learning? (Plag & Winther Balling, 2020). No one has yet provided a rationale for using n-grams to represent aspects of form, while aspects of meaning are encoded only at the whole-word level (cf. Baayen et al., 2019). **Second**, while n-gram models are computationally simple and efficient, it seems highly unlikely that the human processor considers the full range of possible n-grams like a machine (or computer program). **Third**, since n-gram models are usually designed to analyze all kinds of words irrespective of a particular (morphological) task, they follow a one-size-fits-all strategy of processing that creates a combinatorial blowup (and potential problem) for statistical language learning models (Saffran, 2002). **Fourth**, language is a sequential medium in which all elements are arranged in linear order. There is ample evidence that morphological processing is influenced by sequential information (Lelonkiewicz et al., 2020). However, the only distributional information that is explicitly encoded in n-gram models is the occurrence of the initial and final n-grams.

Cognitive scientists who use machine learning to better understand human learning must take all of these concerns into account. What is needed is a task-specific linear processing model that is guided by principles of reliability and efficiency and predicts the emergence of word structure as in natural language use.

In the current study, we present a computational model of morphology that is motivated by psycholinguistic research on sequential processing, task-specificity, and efficiency. Our model is based on network theory (Barabási, 2016) and was introduced by Monakhov and Diessel (2024a, 2024b). In this model, words are represented in a dynamic network of sequentially related sublexical units that emerge from statistical regularities in a large corpus. Monakhov and Diessel used their model to simulate the acquisition and processing of complex English words. In the current study, we use a similar model to analyze the formation of German plural nouns.

The formation of German plural nouns is a good testing ground for the type of model we envisage. The acquisition of the German plural system can be seen as a classification problem whereby novel nouns are assigned to particular plural classes based on their similarity with other plural nouns. Crucially, to evaluate the degree of similarity, one does not need to process the whole word, because for this specific task, word endings are especially informative (Nakisa & Hahn, 1996). Hence, apart from several generic features, such as analogy, linearity, and efficiency, this type of word processing also has one task-specific feature: it is right-oriented. To illustrate, consider the German word *Kraftfahrzeug-Haftpflichtversicherung* 'motor vehicle liability insurance'. To predict the plural marker of this word, an n-gram model will plough its way through evaluating 37 trigrams. However, it is obvious that an n-gram such as *ftf* is less reliable and less informative for this task than the word ending *ung*, which is strongly associated with the plural suffix EN in German (e.g., *Erfahrung-EN* 'experiences', *Zahlung-EN* 'payments', etc.).

In what follows, we show that our task-specific processing model is both computationally more accurate and psychologically more plausible than state-of-the-art n-gram models of the German plural noun system.

Nominal number system in German

In German, plural nouns can take a number of different forms. First, most plural nouns are marked by one of the following five suffixes: N, EN, E, ER, and S. Of these, N and EN are usually analyzed as allomorphs of the same suffix. Second, some plural nouns are morphologically unmarked or zero-coded (Ø), so that number is only indicated by the co-occurring article. Third, many German plural nouns are subject to stem-internal vowel changes (umlaut), which may or may not be used together with a plural suffix (Marcus et al., 1995; Wegener, 1995; 2002; Clahsen, 1999; Bartke et al., 2005). With regard to productivity, (E)N is the most common type of plural suffix, accounting for almost half of all German noun types. Apart from (E)N, E and Ø are fairly common (though Ø is mainly restricted to novel derivations with the agent suffix *-er* and also to nouns ending in *-en*). ER and S occur only with a relatively small number of nouns.

In order to simulate the acquisition of German plural endings, researchers have used various machine-learning techniques, e.g. the nearest-neighbour algorithm (Nakisa & Hahn, 1996), the analogical model (Wulf, 2002), the multilayer perceptron (MacWhinney et al., 1989), the recurrent encoder-decoder (Dankers et al., 2021), the transformer (Beser, 2021), the rule-based decision tree (Belth et al., 2021), and the discriminative learning algorithm (Plag, Heitmeier, & Domahs, 2023). All these models assume an associative learning procedure whereby the model learns to predict the occurrence of a particular plural form from a given singular form.

What we believe to be problematic about these models is that though association learning is based on linguistic cues, the process of their identification remains either unmodelled or modelled in a psycholinguistically implausible way. On the one hand, the models that make use of different grammatical and semantic cues (cf. MacWhinney et al., 1989; Daelemans, 2002; Belth et al., 2021) cannot answer the question of where these cues come from in the first place if they are completely model-external. On the other hand, the models that resort to purely formal, character-based representations (cf. Dankers et al., 2021; Beser, 2021) cannot realistically explain how speakers acquire and store non-symbolic sublexical units serving as pointers to certain grammatical categories. These latter models mostly use ANN architecture and eschew representations for standard linguistic constructs, which makes it difficult to interpret their output in light of specific linguistic hypotheses (Baayen et al., 2019).

The discriminative learning algorithm stands out against other models, first, as based on a cognitively grounded learning theory (Plag, Heitmeier, & Domahs, 2023) and second, as using more concrete representations than ANNs (discriminative learning networks contain no hidden layers and directly map cues to outcomes). However, this approach is also not without its limitations, which we will address in more detail below.

Computational modelling

Data

The data for our study were collected in two steps. First, we retrieved all German noun lemmas together with their token frequencies from the German Web 2020 corpus (Sketch Engine; Jakubíček et al., 2020). As Sketch Engine restricts access to 1,000 items per query, we conducted the search in frequency bands. This allowed us to obtain an exhaustive list of all German nouns with at least 550 tokens in the corpus, rather than a random sample (we stopped at 550 tokens as the number of types exceeded 1,000 for each frequency value). The resulting dataset contained a large number of compounds, which were useless for our purposes as the plural marker of a German compound is determined by its head. Using the database GECOBD (Schäfer & Pankratz, 2018), we replaced all compounds with their heads and removed the duplicates. Nominalized adjectives and participles (e.g., *Jugendliche* 'teenagers'), as well as singularia tantum, pluralia tantum, and proper nouns, were also disregarded.

In the second step, we used the *Digitales Wörterbuch der deutschen Sprache* (*DWDS*) to pair the nouns in our dataset with their respective plural forms. Words not included in the *DWDS* were manually checked in *Duden Online*. All nouns with Latin or Greek plural forms were removed. The resulting dataset comprised 8,079 nouns, represented in phonological forms (i.e. written in IPA symbols), which we obtained from the *DWDS*.

For the purposes of computational modelling, we divided our data into a training (7,079 nouns) and a test (1,000 nouns) dataset. Importantly, since we are interested in plural forms that are learned by structure mapping (or analogy), we selected the 1,000 least frequent plural nouns as our test items (as these forms are unlikely to be rote-learned and listed).

Model

Training To model the type of word processing we envisage, a very simple algorithm was used. The algorithm computed high-confidence and high-efficiency phonological chunks from available input by evaluating empirical probability distributions over plural markers for different segmentation variants. Specifically, each word in the training dataset was analysed as follows. Moving from right to left along the input phonological sequence and increasing the final chunk one phoneme at a time, the algorithm calculated the probabilities of this chunk being associated with each of the plural markers (i.e., respective relative frequencies in the dataset). The process stopped when the probability of any plural marker reached 0.95 (as a conventional significance threshold), after which the respective chunk was selected as a reliable cue that is likely to be identified by most speakers. This is of course a very rough technical approximation of actual language processing. What is important is that the algorithm is guaranteed to maximize the probability of correct decision while minimizing the amount of information to be processed.

Chunk identification was just one part of the problem to be solved. To account for the whole-word processing of German plural nouns, it is necessary to represent these chunks and other parts of word structure in one model. For this, we turned our training dataset into a weighted directed network. The network contained three types of nodes: (i) plural markers, (ii) word-final chunks segmented by our algorithm, and (iii) word-initial chunks obtained by subtraction. A link between any two nodes was added only if this linear sequence of units was attested in at least one lexeme in the dataset. The weights on the links indicated the number of unique words with a particular sequence of adjacent elements. Thus, every word contributed only one unit of weight to each of its constituent links (Figure 1).

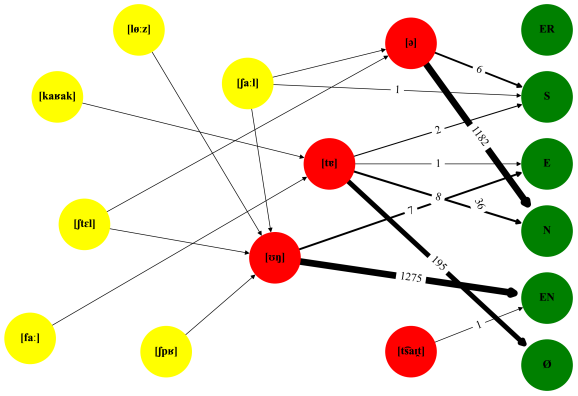


Figure 1: A fragment of the network.

Note that we focused on plural suffixes, disregarding stem changes. As suffixation and umlauting are separate morphological processes (Belth et al., 2021), there are no reasons to believe that they imply the same type of word analysis.

Testing By designing the testing procedure, we used the idea of the shortest path morphological analysis introduced by Monakhov and Diessel (2024a, 2024b). In their approach, any word is conceptualized as a path connecting two or more nodes in a morphological network. As all links in the network are weighted by type frequencies, one can calculate the word's path weight as the sum of the weights of its constituent links. The model is built on the assumptions (i) that among many possible variants of word processing, the one resulting in the shortest path is more likely to be chosen, and (ii) that new words are more likely to be formed along short paths in the network. The concept of shortest path comes from graph theory where it is used to

indicate such a path between two nodes that the sum of the weights of its constituent links is minimized (Schrijver, 2012). To align the type-frequency measure of path weight with the metaphor of shortest path, we used the multiplicative inverses of the weights, which can be conceptualized as distances between adjacent nodes. To illustrate, the weight on the link [ʊŋ] → EN was transformed from 1,275 (cf. Figure 1) to 1 / 1,275 = 0.0007.

To test how useful the identified chunks actually are for learning previously unseen plural nouns, we analysed all nouns in the test dataset against the constructed network. For this, each word to be tested was, first, split into initial and final phonological chunks by the algorithm described above. Importantly, word-final chunks were inferred through evaluating empirical probability distributions over plural markers in the training dataset. Then, the split word was plugged into the network, that is, its nodes and (unity-weighted) links were added to the graph if not already there. The plural marker was selected as the target node reachable from the source node encoding the first chunk of the word by the shortest path. Whenever no path to either plural marker could be found, a missing link was created and weighted by the reciprocal number of links incident upon this node (preferential attachment, Barabási & Albert, 1999). Once the word's plural form was predicted, all newly created nodes and links were deleted, so that all words in the test dataset were analysed against the same state of the network.

Performance The overall accuracy of the model was found to be 98 % for the training dataset, 95 % for the test dataset, and 92 % in leave-one-out cross-validation (Table 1). We interpret the difference in the classification results as an indication of the fact that plural forms of less frequent words are fairly easy to learn by analogy with more frequent words, while learning in the reverse direction is not so beneficial.

Table 1: Accuracy of the computational model.

Marker	Training	Test	LOO-CV
Ø	98 %	97 %	98 %
EN	100 %	96 %	91 %
N	98 %	94 %	95 %
E	98 %	90 %	97 %
S	95 %	76 %	66 %
ER	98 %	—	38 %
Average	98 %	95 %	92 %

Our algorithm outperforms all existing n-gram models of German plural (see *Nominal number system in German*). However, pitting the results in Table 1 directly against those of other studies can be misleading. Some studies include compounds and some do not, some authors consider umlauting and some do not, the number of classification classes can also be very different.

It is interesting to compare our network model to the current state of the art, the discriminative learning model (DL), developed by Baayen and colleagues (see above). Both models are motivated by psycholinguistic research and similar in that they draw on insights from machine learning to better understand the cognitive mechanisms governing morphology.

To compare the performance of the two models, we trained a DL classifier on our training dataset and then used the estimated weight matrix to make predictions about our test dataset. We used the *ndl* R package (Baayen, 2011; Baayen et al., 2011) to compute the association weights between cues (biphones and triphones) and outcomes (plural markers), using the equilibrium equations for the Rescorla-Wagner model of Danks (2003). After estimating the weights, the total activation of an outcome was defined as the sum of the weights on the incoming links from active cues. Probabilities of each outcome were obtained by adding the absolute value of the minimum activation to the activation matrix and renormalizing. For each word in the data, the outcome value with the highest probability estimate was chosen as a predicted plural marker. The results are provided in Table 2, separately for two datasets and two types of cues:

Table 2: Accuracy of the DL classifier.

Cues	Dataset	
	Training	Test
biphones	89 %	87 %
triphones	98 %	82 %

The numbers are similar to those reported by Plag, Heitmeier, and Domahs (2023), who used the same procedure to predict whether a given word form is singular or plural. In Table 2, one can see that the DL model has a low bias and high variance: it overfits the training data and gets confused by multidirectional cues in the test dataset. For example, the word *Macherin* [maxəʁɪn] 'she-maker' takes plural EN, but was mistakenly predicted to take E by the DL model. The main reason for this is that the DL model views the root chunk *mach* [max] as the most important cue, while in reality, the most important cue is the word ending *rin* [ʁɪn] (as correctly recognised by our model).

Our model does not only make more accurate predictions than the DL model but is also more parsimonious. While our algorithm operated on 1,759 word-final chunks, the DL classifier had to deal with 8,378 triphone cues. Importantly, the analysis of these cues reveals very unintuitive patterns. Taken together, word-initial and word-final triphones constitute a minority of all positively weighted cues in the DL matrix. As a positively weighted link between a cue and

an outcome implies that the former is predictive of the latter, one can conclude that most of the heavy lifting in predicting particular plural forms is done by word-internal triphones, many of which appear to be unmotivated, e.g.: [ɔsp] as in *Hospital* 'hospital', [ɪʁb] as in *Wirbel* 'vortex', or [mpç] as in *Lämpchen* 'small lamp'. It is not easy to see what kind of lexical processing can give rise to such units. By contrast, the cognitive reality of word-final chunks identified by our model can be tested in a psycholinguistic experiment.

Psycholinguistic experiment

Design

The purpose of the experiment was to test whether speakers, when finding themselves in a situation where both accuracy of predictions and efficiency of information processing matter, will follow the strategy implemented in our computational algorithm. The experimental setup was as follows. We created a web application (Figure 2) where participants were offered the opportunity to play a simple word game. The goal of the game was to earn as many points as possible while trying to guess a plural marker of a masked German noun.

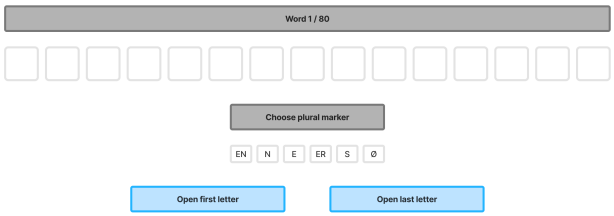


Figure 2: Web application design.

Each word was represented by 15 empty squares, even if it contained fewer letters, so that the participants would not try to guess the word by its length. The players could open as many letters as they wanted, with the only caveat that the letters had to be opened consecutively, one at a time, starting from either the beginning or end of the word. Once a direction was chosen, the two buttons depicted in Figure 2 were automatically replaced with one button saying 'Open one more letter'. The participants submitted their answers by selecting one of the six plural markers. After each submission, an automated message was displayed, indicating (i) whether the participant's answer was true or false, (ii) which word was masked, (iii) how many points the participant scored in a particular round, and (iv) what their points tally was so far in the game.

The maximum number of points participants could gain for each word was 1. Each letter opened reduced this amount by the quantity $1 / N$, where N is the number of letters in the word. The participants gained zero points if either the whole word was opened or they chose a wrong plural marker. Thus, the participants were motivated to open just enough letters, not too many and not too few. The participants were told that they can win a small prize when they achieve a high score in the game.

The game was divided into two phases. In the first phase, the same ten nouns were presented to all participants. In the second phase, each participant worked with their unique set of German nouns. Importantly, at the beginning of the game, participants could freely choose the direction in which they opened letters, but once they had seen the first ten nouns, the button 'Open first letter' disappeared, and moving from right to left became the only option.

30 non-language students of the Friedrich Schiller University Jena, all native speakers of German, took part in the experiment (15 males and 15 females, average age 23 years, SD = 2.85). All participants conducted the experiment on the same laptop in the same room in the presence of a student assistant who made sure that the participants understood the experimental procedure. No subjects dropped out of the experiment. In the second phase, the contributions of two participants were excluded because they tried to guess all plural forms by always opening the same number of letters.

Results

Hypotheses We analysed the results from the first and second phases separately. Our hypotheses were as follows: (i) most participants will initially open letters from left to right (in the reading order), but after a few words they will realise that it is more efficient to open letters starting from the end of words, (ii) most participants will stop opening letters at points that coincide with high-confidence and high-efficiency chunks identified by our algorithm, and (iii) speakers who guess plural markers from these chunks will maximise their chances to respond correctly.

Direction choices Our data for testing the first hypothesis were the direction choices participants made while looking at the first ten words at the beginning of the game. To analyse these data, we fitted a random-intercept Bayesian logistic regression model: $\Pr(y_{j,t} = 1) = \text{logit}^{-1}(\alpha[j] + \beta_1 \times t)$, where $y_{j,t} = 1$ indicates a first-letter choice made by the participant j for the word with index t, $j = 1,\ldots,30$ and $t = 1,\ldots,10$. We used MCMC algorithm to construct the posterior distributions of coefficients and sampled several thousand values from them. The obtained credible intervals and the medians are given in Table 3:

Table 3: Coefficients of the direction-choice model.

Coefficient	5 %	50 %	95 %
intercept	-1.158	-0.263	0.625
word index	-0.333	-0.222	-0.118

As expected, the word index coefficient β_1 was found to be negative. Thus, with every new word, the probability that a participant will start opening letters from left to right decreased by roughly 5 % (-0.222 / 4 = -0.05), with the credible interval [-0.08, -0.02].

Stopping probabilities Our data for testing the second hypothesis were the word-final letter clusters opened by the participants. What we wanted to know is when and why participants decide to select a particular plural suffix. This problem can be couched in terms of survival analysis where the outcome variable of interest is the time until a certain event occurs. In our case, 'time' is measured in terms of opened letters, and 'event' is the choice of a plural marker.

To analyse the data, we fitted a hierarchical random-intercept, random-slope Bayesian model of the form: $\Pr(T_j \geq t) = \exp(-\mathbf{X}_{t,j} \times \boldsymbol{\beta}_{j[w]} \times t^{\alpha})$, where t is the number of opened letters, α and $\lambda = \mathbf{X} \times \boldsymbol{\beta}$ are two parameters of the Weibull survival time probability distribution, $\mathbf{X}$ is a matrix of predictors, and $\boldsymbol{\beta}$ is a vector of coefficients varying by group of words with final chunk j. The stopping parameter λ was modelled as a function of the following predictors: (i) difference in the maximal values of the probability distributions over plural markers associated with clusters j_t and j_{t-1}; (ii) number of word types in the dataset that end with j_t; (iii) summed token frequency of words ending with j_t; (iv) number of word types that contain j_t in any non-final position; (v) summed token frequency of words with non-final j_t; (vi) indicator of whether j_t coincides with any etymological morpheme; (vii) indicator of whether j_t coincides with any reliable phonological chunk; (viii) six indicators for all possible plural markers, and interactions between (ii) and (iv), (iii) and (v), and (vi) and (vii).

We split our experimental stimuli and associated results into the training (clusters with type frequency of at least five words, N = 1,309) and test (clusters with type frequency of less than five and more than one word, N = 247) datasets. The model was fitted to the training data and then used to make predictions about the test data. To save space, only estimates of the coefficients whose confidence intervals do not include zero are provided in Table 4:

Table 4: Coefficients of the stopping-probabilities model.

Coefficient	5 %	50 %	95 %
max.likelihood.diff	0.265	0.745	1.235
final.types	1.264	1.558	1.839
final.tokens	-0.600	-0.380	-0.158
I(chunk)	0.146	0.472	0.795
…	…	…	…

The coefficients in Table 4 have a clear interpretation. For a given number of opened letters t and a fixed value of α, the survival probability decreases with increasing value of λ. Diminishing survival probability, of course, implies growing probability of 'death'. Hence, positive coefficients indicate speakers' greater confidence in the unmasked clusters, that is, their readiness to stop opening letters and to guess a plural marker. Negative coefficients, by contrast, can be taken as a sign of speakers' greater uncertainty.

From Table 4, one can draw the following conclusions. First, if a certain marker becomes much more probable than any other marker considered previously, speakers tend to select it. Second, if the cluster is such that speakers encounter it in many word types, they identify it as a reliable cue. Third, if the cluster is a familiar ending of several very frequent words, but the overall number of words with it is average, speakers find such evidence not supportive enough to make a final decision. Finally, chunks identified by our algorithm are characterised by their high recognisability in noun pluralisation tasks.

Correct answers The final issue we investigated was whether speakers who based their decisions on reliable chunks maximised the chances of correct answer, as a model-free estimation of conditional probabilities suggests: $\Pr(guess = True \mid cluster = chunk) = 0.74$, $\Pr(guess = True \mid cluster \neq chunk) = 0.60$. To test our third hypothesis, we fitted a Bayesian binomial regression model of the form: $\Pr(\Theta = \theta_i) = \text{logit}^{-1}(\mathbf{X}_i \times \boldsymbol{\beta})$, where θ_i is the probability of success for the opened cluster i, and $\mathbf{X}$ is a matrix of the same predictors as introduced in *Stopping probabilities* minus maximum likelihood difference and plus cluster length in letters.

Overall, the results were similar to those reported for the survival analysis model. The probability of success θ_i was found to be positively influenced, for each opened cluster, by (i) its length, (ii) the summed token frequency of words ending with it, and (iii) the indicator of whether it coincides with any phonological chunk. The first finding is self-explanatory: the more letters are opened, the easier it is to guess the word and to map it onto a particular plural form. The second finding shows that clusters belonging to very frequent words provide better cues for a particular plural class than clusters appearing in infrequent, albeit numerous words. The third finding supports the hypothesis that the chunks identified by our algorithm are characterised by their reliability in solving the task of selecting correct German plural suffixes.

Conclusion

In this paper, we argued that traditional n-gram models of word comprehension and production suffer from several weaknesses: (i) they are computationally redundant (ii) and cognitively implausible (or difficult to interpret), and (iii) they pay little attention to the sequential organization of words. As an alternative, we proposed a computational network model that implements a task-specific linear processing algorithm guided by general principles of efficiency and reliability. We showed that our model does not only make accurate predictions about the acquisition of German plural forms but also appears cognitively plausible, because the sublexical units we identified were found to match those recognised by speakers in a psycholinguistic experiment.

References

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology 3*: 1–12.

Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics 11*: 295–328.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (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 LDL. *Complexity*: 4895891.

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

Barabási, A.-L. (2016). *Network Science*. Cambridge: Cambridge University Press.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science 286 (5439)*: 509–512.

Bartke, S., Rösler, F., Streb, J., & Wiese, R. (2005). An ERP-study of German 'irregular' morphology. *Journal of Neurolinguistics 18(1)*: 29–55.

Belth, C., Payne, S., Beser, D., Kodner, J., & Yang, C. (2021). The greedy and recursive search for morphological productivity. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society 42(1)*: 2869–2875.

Beser, D. (2021). Falling through the gaps: Neural architectures as models of morphological rule learning. *arXiv*: 2105.03710.

Blevins, J. P. (2016). *Words and paradigms*. Oxford, England: Oxford University Press.

Booij, G. E. (2010). C*onstruction morphology*. Oxford: Oxford University Press.

Booij, G. E. (2016). Construction morphology. In A. Hippisley, & G. T. Stump (Eds.), *The Cambridge handbook of morphology*. Cambridge: Cambridge University Press, pp. 424–448.

Bybee, J. (1985). *Morphology. A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes 10(5)*: 425–455.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Chuang, Y.-Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2021). The processing of pseudoword form and meaning in production and comprehension. *Behavior Research Methods 53*: 945–976.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences 22(6)*: 991–1013.

Daelemans, W. (2002). A comparison of analogical modeling to memory-based language processing. In R. Skousen, D. Lonsdale, D. B. Parkinson (Eds.), *Analogical Modeling: An exemplar-based approach to language (Human Cognitive Processing. Vol. 10)*. Amsterdam/ Philadelphia: John Benjamins, pp. 157–179.

Dankers, V., Langedijk, A., McCurdy, K., Williams, A., & Hupkes, D. (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*. Association for Computational Linguistics, pp. 94–108.

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

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

Grandon, B., Schlechtweg, M., & Ruigendijk, E. (2023). Processing of noun plural marking in German-speaking children: an eye-tracking study. *Journal of Child Language*: 1–28.

Jackendoff, R., & Audring, J. (2020). Morphology and memory: Toward an integrated theory. *Topics in Cognitive Science 12(1)*: 170–196.

Jakubíček, M., Kilgariff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2020). *German Web 2020 (deTenTen20)*. https://www.sketchengine.eu/detenten-german-corpus/ #toggle-id-2. Accessed on October 28, 2024.

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.

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

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

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language 115*: 104152.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum, pp. 249–308.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language 28(3)*: 255–277.

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review 126(1)*: 1–51.

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

Matthews, P. (1991). *Morphology* (2nd ed.). Cambridge: Cambridge University Press.

Milin, P., Smolka, E., & Feldman, L. B. (2017). Models of lexical access and morphological processing. In: E. M. Fernandez & H. Smith Cairns (Eds.), *The Handbook of Psycholinguistics*. Blackwell Handbooks in Linguistics. Wiley-Blackwell, pp. 249–268.

Monakhov, S., & Diessel, H. (2024a). Complex words as shortest paths in the network of lexical knowledge. *Cognitive Science 48(11)*: e70005.

Monakhov, S., & Diessel, H. (2024b). A network model of English derivational morphology. In *Proceedings of the Annual Meeting of the Cognitive Science Society 46*, pp. 4732–4738.

Mugdan, J. (1977). *Flexionsmorphologie und Psycholinguistik*. Tübingen: Narr.

Nakisa, R. C., & Hahn, U. (1996). Where defaults don't help: the case of the German plural system. In G. W. Cotrell (Ed.), P*roceedings of the Eighteenth Annual Conference of the Cognitive Science Societ*y. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 177–182.

Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology 34*: 393–418.

Pescuma, V. N., Zanini, C., Crepaldi, D., & Franzon, F. (2021). Form and function: A study on the distribution of the inflectional endings in Italian nouns and adjectives. *Frontiers in Psychology 12*: 720228.

Plag, I., Heitmeier, M., & Domahs, F. (2023). German nominal number interpretation in an impaired mental lexicon: A naive discriminative learning perspective. *The Mental Lexicon 18(3)*: 1–25.

Plag, I. & Winther Balling, L. (2020). Derivational morphology: An integrative perspective on some fundamental questions. In *Word knowledge and word usage*. Berlin, pp. 295–335.

Plaut, D., & Gonnerman, L. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes 15(4/5)*: 445–485.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review 11(6)*: 1090–1098.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Lawrence Erlbaum Associates, Inc., pp. 195–248.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language 44*: 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. J*ournal of Memory and Language 47*: 172–196.

Schäfer, R., & Pankratz, E. (2018). The plural interpretability of German linking elements. *Morphology 28(4)*: 325–358.

Schrijver, A. (2012). On the history of the shortest path problem. In M. Grötschel (Ed.), *Documenta mathematica. Extra volume: Optimization stories. 21st international symposium on mathematical programming*. Washington: SPARC Leading Edge, pp. 155–167.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science 14*: 238–242.

Skousen, R., Lonsdale, D., Parkinson, D. B. (Eds.) (2002). *Analogical Modeling: An exemplar-based approach to language* (Human Cognitive Processing. Vol. 10). Amsterdam/Philadelphia: John Benjamins.

Smolka, E., Libben, G., & Dressler, W. U. (2019). When morphological structure overrides meaning: Evidence from German prefix and particle verbs. *Language, Cognition and Neuroscience 34(5)*: 599–614.

Smolka, E., Preller, K., & Eulitz, C. (2014). 'Verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language 72(1)*: 16–36.

Wegener, H. (1995). *Die Nominalflexion des Deutschen—verstanden als Lerngegenstand*. Tübingen.

Wegener, H. (2002). Aufbau von markierten Pluralklassen im Deutschen—eine Herausforderung für die Markiertheitstheorie. *Folia Linguistica 36(3–4)*: 261–295.

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* (Human Cognitive Processing. Vol. 10). Amsterdam/Philadelphia: John Benjamins, pp. 109–122.