Open Library of Humanities

Modelling German Word Stress

Fabian Tomaschek, Seminar für Allgemeine Sprachwissenschaft, Eberhard Karls University of Tübingen, Germany, fabian.tomaschek@uni-tuebingen.de

Ulrike Domahs, Institut für Germanistische Sprachwissenschaft, Philipps-Universität Marburg, Germany, domahsu@staff.uni-marburg.de

Frank Domahs, Seminar für Sprachwissenschaft, University of Erfurt, Germany, frank.domahs@uni-erfurt.de

Standard linguistic and psycholinguistic approaches to stress assignment argue that the position of word stress is determined on the basis of abstract information such as syllable weight and number of syllables in the word. In the present study, we contrasted this approach with a perspective based on learning analogies according to which speakers learn to associate word form cues to stress position. To do so, we use a simple two-layer neural network trained with an error-driven learning mechanism to predict stress position in German morphologically simple and complex words. We find that networks trained on word forms outperformed networks trained on cues that represent abstract information. Moreover, most standard approaches assign stress from right to left. We tested this proposal and found that in morphologically simple words, assignment from right yielded better results than assignment from left, supporting the standard approach. By contrast, in morphologically complex words assignment from left outperformed assignment from right. We discuss the implications of our results for psycholinguistic theories of stress assignment by taking into account word form cues, abstract cues, assignment direction, and the representation of stress in the mental lexicon.

1 Introduction

In languages with free stress, like German, the question arises what factors determine the position of stress in a word. In the present study, we investigate this question by testing how different kinds of cues provided to a simple two-layer neural network determine the network's accuracy in predicting stress position in German morphologically simple and complex words. We test an approach, according to which stress is assigned on the basis of similarity of word forms, where the similarity is learned through an error-driven learning process. We contrast this similarity-based approach with an approach using cues that represent more abstract types of information such as number of syllables or syllable structure that are typically considered in more traditional linguistic and psycholinguistic approaches.[1] We find that assignment based on similarity outperforms assignment based on cues that represent abstract information. Moreover, we find that in morphological simple forms, assignment from right yields better results than assignment from left, supporting the standard approach. By contrast, in morphologically complex words assignment from left outperforms assignment from right. When analyzing how the network represents knowledge about stress assignment, we demonstrate that individual cues are not capable to predict a specific stress position when they are considered in isolation. Only in context with other cues, they demonstrate their predictive power about a stress position. In the remainder of this introduction, we present the theoretical and empirical background on word stress assignment in German. Subsequently, we discuss the material used in the present study as well as the computational method to investigate stress assignment. After presenting the results of our computational experiments, we discuss the implications of our results for psycholinguistic theories of stress assignment by taking into account word form cues, abstract cues, counting direction, and the representation of stress in the mental lexicon.

1.1 Word stress in German

1.1.1 Syllable weight

Several factors have been proposed to play a role in stress assignment. According to parametric accounts (e.g., Hayes 1995) and constraint-based accounts (e.g., Kager 1999) of word stress assignment, parameter settings or constraint rankings can be defined based on properties like edge-marking (left- vs. rightmost), number of feet (one vs. more than one), foot-type (trochaic vs. iambic), parsing direction (left- vs. rightwards), and quantity-sensitivity (Féry 1998; Jessen 1999). However, the suggested analyses of where to place stress in German are controversial. Apart from accounts assuming that word stress has to be lexicalized as it is largely idiosyncratic

---

[1] We are aware of the fact that phonemes, graphemes and syllables present some kind of abstract linguistic units. However, as the same given syllable structure (or CV-skeleton) represents different phoneme/grapheme sequences, we consider specific sequences of phonemes/graphemes as less abstract than constructs like "number of syllables" or "syllable structure". For convenience, we will use the term "abstract" to refer to the latter type of constructs.

(Levelt et al. 1999), proposals differ basically according to the question whether syllable weight influences stress assignment or not. Following theories neglecting syllable weight, stress is assigned to the initial or penultimate syllable by default or – in cases deviating from the default – specified lexically to occur on the non-initial or final / antepenultimate syllable (Levelt et al. 1999; Wiese 2000). By contrast, in quantity sensitive accounts, not only the syllable structure of the rhyme, but also the position of syllables with a certain structure seem to play a role in stress assignment (Giegerich 1985; Vennemann 1991; Féry 1998). For example, Domahs et al. (2014b) analyzed the probability of stress positions in German trisyllabic words taken from the CELEX lexical database (Baayen et al. 1993) and found that words with a heavy final syllable are likely to receive either final stress or antepenultimate stress if the penultimate syllable is light (σLˈH or ˈσLH). In words with a light final syllable, the penult is most likely stressed, even more so if the penult itself is heavy (ˈLL or ˈHL). Syllable structure of the antepenultimate syllable is less commonly discussed as relevant, but there is evidence for a weak influence of antepenultimate rhyme structure on the stress pattern of German words (Röttger et al. 2012).

The observation that the word stress position depends on the structure of the final syllable is interpreted as evidence for the existence of trochaic foot structure in German (Giegerich 1985; Féry 1998; Janssen 2003). A heavy final syllable can be parsed as a monosyllabic foot and additional syllables to the left as additional bisyllabic feet (e.g., *Vitamin* [(vi.ta)$_F$.(ˈmiːn)$_F$] 'vitamin'). This way, trisyllabic words ending in a heavy syllable can be parsed into two feet, while those with a light final syllable only into one (e.g., *Zitrone* [tsi.(ˈtroː.nə)$_F$]) 'lemon'). In addition, also the number of syllables within a word and their parity (even or odd) may affect the parsing of syllables into feet (Ernestus & Neijt 2008). Words with an even number of syllables tend to build bisyllabic feet where the rightmost foot receives main stress (i.e. penultimate stress), while words with an odd number of syllables tend to receive final or antepenultimate stress (Janßen & Domahs 2008). In contrast to the rhyme structure, onset complexity seems to have only a small influence at best (Mengel 2000; Röttger et al. 2012).

1.1.2 Non-phonological factors in stress assignment

Most accounts to stress assignment consider phonological properties of spoken language. However, other lexical aspects have also been shown to affect stress assignment. For example, when it comes to reading, orthographic properties become relevant as well. While, in alphabetic scripts, orthographic structure typically reflects phonological structure, this relation may not always be free from ambiguity. Using complex graphemes which code simple phonemes, Röttger et al. (2012) have demonstrated that orthographic syllable weight can influence stress assignment in German independently from phonological syllable weight. In consequence, models for stress assignment should be specific on the modality they are operating on (spoken vs. written language).

Another factor that very likely affects stress assignment is part of speech. In English for example, initial stress is more often assigned to nouns, whereas disyllabic verbs tend to bear stress on the final syllable (e.g. Davis & Kelly 1997). For German, too, stress patterns have been described to depend on lexical class. Eisenberg (1991), for instance, considers stress patterns of word forms organized in paradigms rather than patterns of stems and suggests a trochaic word pattern at the end of inflected nouns (e.g., *Magazine* [ma.ga.'tsi:.nə] 'magazines'), while inflected adjectives often end in a dactylic pattern (e.g., *größere* ['gʁø:.sə.ʁə] 'bigger SG-FEM').

Furthermore, stress position may also be influenced by morphological structure (Giegerich 1985; Wiese 2000; Eisenberg 2016; Alber & Arndt-Lappe 2020). In the framework of lexical phonology, the stress position depending on suffixes has led to a crucial distinction between stress attracting and stress neutral suffixation (Giegerich 1985; Wiese 2000; Eisenberg 2016). This distinction was supported in reading experiments with suffixed pseudowords for which it has been found that suffixes may either attract stress or modify stress positions (Rastle & Coltheart 2000). Both simplex and derived forms have been argued to obey the three-syllable window, meaning that word stress is assigned to one of the three final syllables of words (Vennemann 1991; Jessen 1999). In contrast, however, compound words are not restricted by the three-syllable window. Since they are often stressed on the initial constituent, main stress is placed outside of the three-syllable window whenever polysyllabic words are combined (Wiese 2000).

1.1.3 Direction of stress assignment

Another question strongly debated in the field of stress assignment concerns the direction of assignment. Does stress assignment operate from left to right or the other way around? Given that, within a word, the structure of the final and prefinal syllables are particularly relevant for stress assignment (Mengel 2000; Röttger et al. 2012; Domahs et al. 2014b), processing from right to left (i.e., starting at the word's end) seems most expedient for stress assignment. Moreover, the end of the word may be particularly informative for establishing analogies (Burani et al. 2014) and for analyzing stress-relevant morphological suffixes (Rastle & Coltheart 2000). Additionally, the three-syllable window highlights the particular relevance of a word's right edge in stress assignment. In fact, a psycholinguistic study investigating the direction of stress assignment via its correlation with working memory supports the proposal that stress assignment proceeds from right to left (Domahs et al. 2014a). However, there are also alternative accounts that argue for stress assignment taking place from left to right (e.g. Levelt et al. 1999; Mattys & Samuel 2000; Schiller et al. 2006). This assumption is predominantly motivated by the fact that a huge amount of native words in German is bisyllabic, ending in an (unstressable) schwa syllable (Eisenberg

1991; Féry 1998; Alber 2020). Indeed, from a historical perspective, German can be described as a language with fixed stress on the first syllable until the 16th or 17th century (Speyer 2009).

1.1.4 The role of analogy

A further factor to be considered is the role of analogy which has been demonstrated in a number of languages including Italian (Burani et al. 2014) and English (Guion et al. 2003; Arciuli et al. 2010; Moore-Cantwell 2020). According to the analogical approach, stress is assigned on the basis of similarity between word forms. Domahs et al. (2014b) demonstrated this for English, where words ending in -y typically have stress on the antepenult syllable. Specifically, sets of words sharing an ending and a stress pattern (i.e., 'friends') seem to be a relevant factor in assigning stress to words in reading Italian (Burani et al. 2014). A similar process may also be assumed in German.

1.2 The present study

As we have argued above, the allocation of word stress in German seems to involve both phonological and morphological properties that can be considered best if stress is assigned from right to left. However, these assumptions are not uncontroversial.

First, in some words phonological and morphological principles may conflict with each other. Since directionality interacts likewise with syllable structure and morphological structure, we will test the influence of directionality on stress assignment. Second, most discussed accounts of stress assignment are based on some kind of abstract structures that are defined in a top-down fashion. Nevertheless, word stress data are not fully explainable by structural phonological rules (Giegerich 1985; Daelemans et al. 1994; Van Oostendorp 2012; Domahs et al. 2014b). Accordingly, the current study follows a naïve approach that is deprived of rules and abstract suprasegmental units, but instead focuses on stress assignment on the basis of phonological similarity among word forms. Finally, most empirical accounts of German word stress are based on morphologically simple disyllabic or trisyllabic words (e.g. Janssen 2003; Röttger et al. 2012; Domahs et al. 2014b) that constitute a small subset of the lexicon (as we will demonstrate below in **Figure 2**). Thus, morphologically complex words are typically ignored when investigating the position of primary stress. The current study also takes this part of the lexicon into account and investigates to what degree stress assignment varies in morphological simple words and morphological complex words.

In light of these problems and given the many different theoretical accounts, we approach the question of word stress assignment from a naïve perspective. We follow the approach by Arndt-Lappe et al. (2022) and train a simple, yet powerful two-layer neural network, the Naïve Discriminative Learner (NDL, Baayen et al. 2011; Arppe et al. 2018) (more details on the network below in the Methods section). Crucially, studies favoring rule based accounts fail to explain how

these rules are actually learned. The use of NDL may provide an answer to this question, since it is trained with a learning mechanism that has been demonstrated to be cognitively valid, mirroring human learning behavior (Ramscar et al. 2010, 2013a; Ramscar 2021). Accordingly, the present study proposes a model of how stress assignment might be acquired.

To predict stress position, the network is provided with relatively unstructured orthographic and phonological cues, which do not contain any information related to suprasegmental structures as discussed above. Then, we test to what degree the trained network is capable to predict the stress position for a given word (which we regard as the assignment process), when it is provided with these specific cues in contrast to more abstract cues. For this purpose, we consider various features and factors that previous studies have argued to determine word stress positions – the direction of stress assignment (leftward vs. rightward), structure (heavy vs. light), and number of syllables (both, absolute number and parity), linguistic modality (phonological vs. orthographic), word class (noun, adjective, verb, function word), as well as morphological structure (simplex vs. complex forms). In sum, we want to investigate what cues and factors are relevant that the network is capable to successfully predict the notoriously complex stress assignment in German.

2 Methods

2.1 Material

A total of 85,495 lemma types with a frequency count of at least 1 was selected from the German CELEX corpus (Baayen et al. 1993) to serve as training and test material in the present study. The morphologically simple and complex word forms that served as the study's material consisted of 35,733 nouns, 26,297 verbs and verbal participles, 23,335 adjectives and 130 articles. Word forms were tagged with information about syllable structure and part of speech – both types of information were provided in CELEX. German word forms are often homophones such that they represent multiple inflected forms (i.e., syncretisms). To account for these syncretisms, we created one entry for each inflected form. In this way, the total number of inflected word forms used to train the network amounted to 199,626 lexeme tokens.

The modelling approach in the present study follows the approach in Arndt-Lappe et al. (2022) who used the same two-layer neural network as the present study to classify stress position in English on the basis of orthographic bigram and trigram cues. Stress position was counted either from the onset of the word (stress from left) or offset of the word (stress from right). Accordingly, we use these two types of counting directions as outcomes in the present study, representing stress assignment processes in different psycholinguistic theories (Levelt 1999; Schiller et al. 2006; Domahs et al. 2014a). **Figure 1** (left and mid) illustrates the frequencies of stress positions depending on the counting direction.
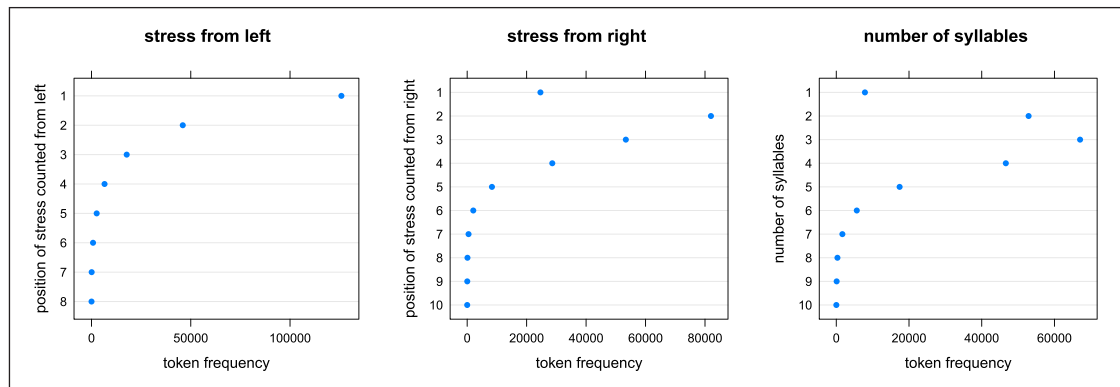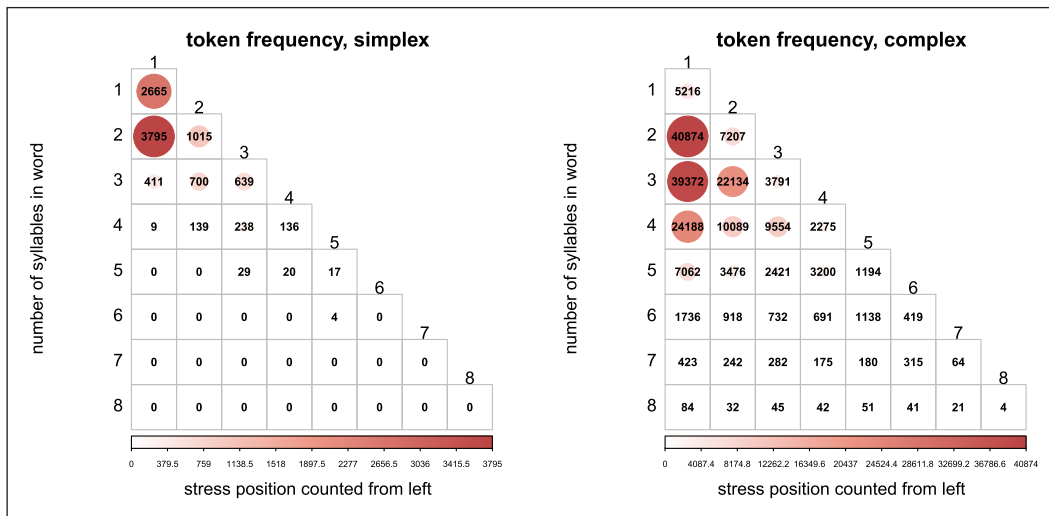
Figure 1: Left and mid: Frequency of occurrence of stress positions depending on counting direction. Right: Type frequency depending on number of syllables.

Like in Arndt-Lappe et al. (2022), we used phonological and orthographic bigrams and trigrams as cues in the present study. However, linguistic and psycholinguistic research has repeatedly shown that information from higher levels of abstraction may also be relevant for the stress position in German. These pieces of information include number of syllables and parity (Alber 1998; Janßen & Domahs 2008), syllable structure (Giegerich 1985; Alber 1998; Féry 1998; Domahs et al. 2008; Janßen & Domahs 2008; Röttger et al. 2012, see Turk et al. 1995 for English), and part of speech (Eisenberg 1991, 2016). Accordingly, we also tested how the network's classification accuracy differed when these pieces of information were used as cues on their own as well as jointly during training.

In the corpus, word length measured by the number of syllables ranged between one and ten. **Figure 1** (right) illustrates their frequency distribution. Trisyllabic word tokens were most frequent in the corpus, followed by disyllabic and tetrasyllabic ones. Notice the relatively small number of monosyllabic word tokens. The number of different syllable structure combinations including syllables with ambisyllabic consonants amounted to 58 (not illustrated). The four different word class categories summed up to 163 different types of inflectional categories, once grammatical information such as person, number, mode and tense for verbs, number and case for nouns and articles, and number, case and positive, comparative and superlative forms for adjectives were taken into account.

Figure 2 illustrates how many stress tokens our data set contains by taking into account the stress position and the number of syllables in a word. The left panel focuses on morphologically simple words, the right on morphologically complex words. Clearly, morphologically simple words are in the minority and basically restricted to a word length of one to four syllables: Our data contained 9,817 morphologically simple, in comparison to 189,809 morphologically complex words.

token frequency, simplex

number of syllables in word	1	2	3	4	5	6	7	8
1	2665							
2	3795	1015						
3	411	700	639					
4	9	139	238	136				
5	0	0	29	20	17			
6	0	0	0	0	4	0		
7	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0

stress position counted from left

token frequency, complex

number of syllables in word	1	2	3	4	5	6	7	8
1	5216							
2	40874	7207						
3	39372	22134	3791					
4	24188	10089	9554	2275				
5	7062	3476	2421	3200	1194			
6	1736	918	732	691	1138	419		
7	423	242	282	175	180	315	64	
8	84	32	45	42	51	41	21	4

stress position counted from left

Figure 2: Number of tokens depending on word length (in syllables) and stress position counted from left in morphologically simple (left panel) and complex words (right panel).

A note on morphological complexity. Upon checking the information about morphological complexity provided by the CELEX corpus, we found that a larger number of morphologically complex words was wrongly classified as morphologically simple. This is why the information provided by CELEX was replaced by information which we constructed on our own. Using an automatic process, we checked whether mono-, bi-, and trisyllabic words were present in words with a higher number of syllables. Through this procedure, the majority of words in our corpus was classified as "morphologically complex". We manually checked a sample 500 words classified as morphologically complex and found that only eleven were misclassified (i.e. 2.2%). In other words, we regard our approach to deliver a more reliable classification of morphological complexity than the information provided by CELEX. We understand that there may be some words that have been incorrectly classified. However, given the large number of morphologically complex words, we did not manually correct all of them. By contrast, we wanted to ensure that the list of morphologically simple words was correct. This is why the subset of morphological simple words was manually corrected. To this end, all verbs, including the infinitive forms, were counted as morphologically complex.

2.2 Modelling approach

We use the Naïve Discriminative Learner (NDL, Arppe et al. 2018; Baayen et al. 2011) to model to what degree orthographic and phonological cues are predictive of stress position. NDL presents a simple two-layer neural network with one input and one output layer. We used the Danks Equilibrium Equations (Danks 2003) provided by the NDL package to calculate the connection weights between the input and the output layer. The Danks Equilibrium Equations allow a fast

computation of connection weights between cues and outcomes by estimating a state when training using the error-driven learning equations by Rescorla and Wagner (Rescorla & Wagner 1972) reaches an equilibrium.

The functionality of the two algorithms has been successfully demonstrated to capture discriminative learning (Bröker & Ramscar 2020; Ramscar 2021) of morphological processes including inflection (Ramscar & Yarlett 2007; Ramscar et al. 2013b, 2010; Nieder et al. 2021, 2022), of processing in the context of reading and listening to morphological simple and complex words (Baayen et al. 2011; Arnold et al. 2017), and also in other domains such as the learning of phonetic categories (Olejarczuk et al. 2018; Nixon 2020; Nixon & Tomaschek 2020, 2021) and speech production (Ramscar & Yarlett 2007; Tomaschek et al. 2019; Baayen et al. 2019; Tomaschek & Ramscar 2022). An introduction to using NDL can be found in Tomaschek (2020) and an excellent overview of how the dynamics of connection weights depend on cue-to-outcome constellations is presented by Hoppe et al. (2022).

Since these studies have extensively explained the mathematical details of the Danks and Rescorla-Wagner equations, we present their functionality in a nutshell. These algorithms learn to associate cues to outcomes through prediction and prediction-error (which is the case for all learning algorithms that take into account some kind of error). This means that changes to the connection weights during training are proportional to the difference between prediction strength of the predicted and the experienced outcome – with the difference depending on the amount of phonological/orthographic overlap between these two instances. Moreover, connection weights are also adjusted on the basis of non-occurrences between cues and outcomes. Since the error-driven learning algorithm adjusts connection weights on the basis of occurrences and non-occurrences between cues and outcomes, it indirectly takes into account similarities among words and how these relate, in the present case, to stress. In this way, the algorithm is able to represent stress assignment on the basis of phonological/orthographic neighbors, as has been demonstrated in previous studies (e.g. Guion et al. 2003; Arciuli et al. 2010; Burani et al. 2014; Moore-Cantwell 2020).

Training our network can be considered to be equivalent to the learning task that speakers face when learning how to associate word forms to a stress position. Testing our model is equivalent to an experiment during which speakers/readers have to pronounce words and stress them appropriately. Specifically, our test operationalizes the moment during cognitive preparation of speech, when the stress position is selected on the basis of the information presented by the word form.

2.3 Cue-to-outcome structure

To demonstrate how we constructed our cues and outcomes for the learning networks, take the word *Modell* 'model'. This disyllabic word has as outcomes stress from left = 2 and stress from right = 1. **Table 1** illustrates the cue types for the word.

type of cue	cue
inflectional information	noun, nominative, singular
number of syllables	2
parity	even
syllable structure	#.CV, CV.CVC, CVC.#
orthographic bigrams	< #M, Mo, od, de, el, ll, l# >
orthographic trigrams	< #Mo, Mod, ode, del, ell, ll# >
phonological bigrams	[#m, mo, od, dɛ, ɛl, l#]
phonological trigrams	[#mo, mod, odɛ, dɛl, ɛl#]

Table 1: Cue structure for the word 'Modell' (Engl. *model*), which has as outcomes stress from left = 2 and stress from right = 1. Syllable structure was included as bisyllabic sequences where each syllable is separated by a dot. The hashtag represents the word edge.

The upper part illustrates the cue types containing abstract higher level information (inflectional information, number of syllables, syllable structure). We constructed a power set of these cue types. The subsets of the power set are permutations of one type, two types, three types and all four types of cues, including an empty set. All these sets were trained either on their own, or combined with one type of n-gram cues, illustrated in the lower part of **Table 1**. These could be either orthographic bigrams, orthographic trigrams, phonological bigrams or phonological trigrams.

Combining different types of n-gram cues with different types of abstract cues amounted to 80 different cue type combinations to be tested. Since we also tested counting stress position from left and from right as an outcome, the total number of cue structures used to train and test the network amounted to 158.

2.4 Testing trained networks

Training the network modulated the association weights between cues and outcomes. Cue-outcome-combinations in which the cues are informative about the outcome obtain positive weights; cues that are uninformative about an outcome obtain negative weights. Once a network was trained on the presented input, we used it as a classifier to predict stress position. This was accomplished by presenting the network with a set of cues, such as those in **Table 1**. The weights between the presented cues and all possible outcomes were summed resulting in an *activation value* which operationalizes how strongly the set of cues supports a stress position outcome. The stress position outcome with the highest activation was selected as the winner of the classification for a specific cue set. In this way, we obtained a stress position predicted by the network. This predicted stress position was compared to the stress position provided by the CELEX corpus. A

match between the predicted and CELEX stress position was coded as CORRECT, a mismatch was coded as INCORRECT. We used this coding to assess the accuracy of the networks by counting the number of correctly predicted stress positions. In the next section, we will discuss the performance of the networks trained based on different cue-to-outcome combinations.

3 Results

3.1 Evaluating network performance

The accuracy for all networks ranges between 34.7% and 91.8%, measured as the number of correctly identified stress positions divided by the number of individual word-forms in the data set. The high upper bound indicates that, given the appropriate cue structure, stress position can be reliably predicted without any type of rules. However, to evaluate network performance, accuracy is inadequate because it is based only on true positives, but misses false positives and false negatives. This shortcoming is alleviated by the F-score which is typically used in computational modelling studies. It is calculated by combining the measures 'precision' and 'recall'. Precision takes into account the number of true positives and false positives and as such "measures the percentage of system-provided [items] that were correct" (Jurafsky 2000, p. 489). Recall takes into account the number of true positives and false negatives and "measures the percentage of [items] actually in the input that were correctly identified" (ibd.). F-scores are bound between 0 and 1, with 0 representing a network that completely failed in the classification task and 1 representing a network that is excellent at classification.[2]

To gauge network performance, we calculated *weighted average F-scores,* by multiplying the F-scores for each stress position by the number of items with that specific stress position and dividing the summed F-score by the total amount of items. In this way, stress positions with many items (e.g. stress from left = 1) weigh heavier than stress positions with only a few items (e.g. stress from left = 8) in the estimation of the weighted average F-score.

In the following paragraphs, we used beta regression (package betareg, Version 3.1–4 Cribari-Neto & Zeileis 2010) to test differences in F-scores. Beta regression allows the researcher to model dependent variables bound between 0 and 1.[3] To do so, beta regression transforms the values ranging between 0 and 1 into logits and calculates the difference between factor levels in logits.

Whenever we report the results of beta-regression below, we used a model structure in which F-scores were fitted as a function of the variable of interest. All figures below that illustrate

---

[2] Precision is calculated by dividing the number of true positives by the sum of the number of true positives and false positives. Recall is calculated by dividing the number of true positives by the sum of the number of true positives and false negatives. F-score is calculated by dividing the two times product of precision and recall by the sum of precision and recall.

[3] Linear regression could not be performed here because it assumes that the dependent variable can theoretically range between plus/minus infinity.

F-scores were obtained by averaging F-scores for the combination between variables plotted on the x-axis and the variables illustrated by means of line-type.

Figure 3 illustrates weighted average F-scores as a result of different kinds of cue-combinations. Overall, we find that counting the stress position from left, i.e. from word onset, yields a better accuracy than counting from right, i.e. from word offset ($\beta = 0.75328$, $sde = 0.08993$, $z = 8.377$, $p < 0.0001$). The result makes sense if we consider that the corpus contains a large amount of suffixed words. Suffixation in German typically changes a word's number of syllables, while in many cases the stressed syllable remains the same. Accordingly, counting the stress position from right is reflected by greater uncertainty about the stress position than counting from left.
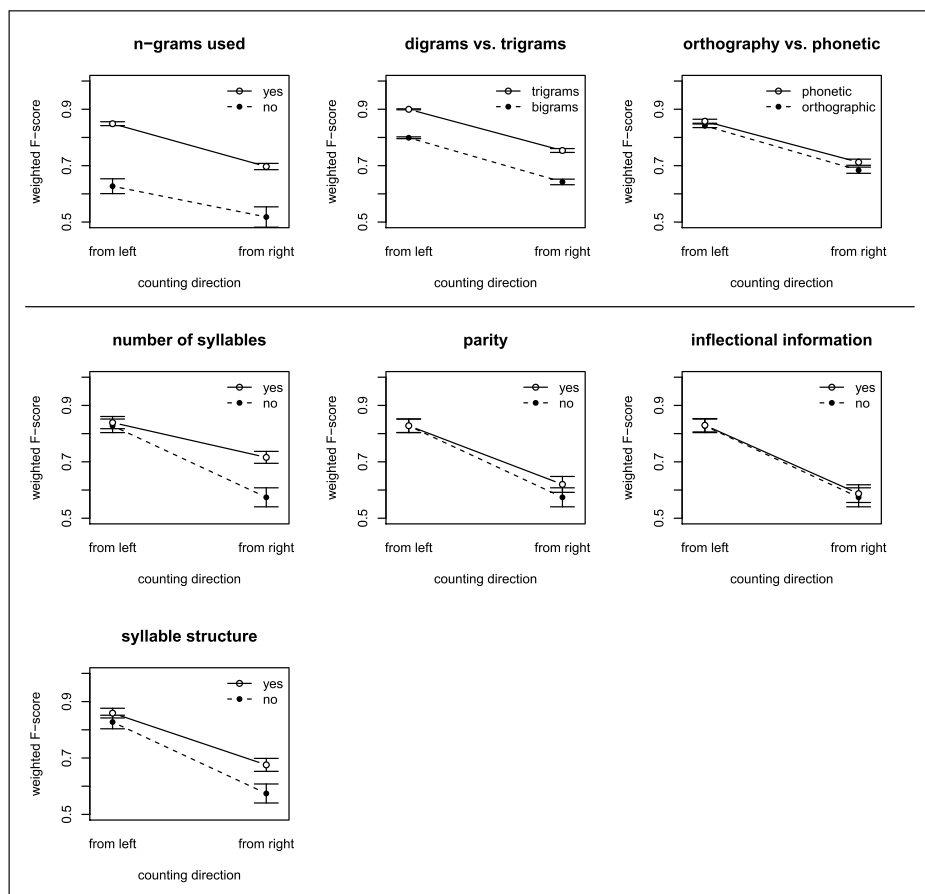
Figure 3: Average weighted F-scores for all 158 networks tested with the training material. The x-axis illustrates the results depending on counting direction. The plots in the top panel (above the line) illustrate changes in F-score depending on whether n-grams were used (left) and n-gram type (mid and right). The plots in the mid and bottom row below the line illustrate the results from models which always contained n-grams as cues in addition to the presence or absence of the cue indicated in the title: Thus, they contrast to what degree the addition of number of syllables, parity, syllable structure and inflectional information as cues increased F-scores.

Turning our attention to the type of cues used, we find that the network is better when it was trained with abstract cues in combination with n-grams than when it was trained only with abstract cues (**Figure 3**, top left, $\beta = 0.93434$, $sd = 0.10501$, $z = 8.897$, $p < 0.0001$). We observe that using trigrams improves network performance in contrast to using bigrams ($\beta = 0.6195$, $sd = 0.0624$, $= 9.928$, $p < 0.0001$). Note that using trigrams when counting stress from left yields the best network accuracy of roughly 90%. Furthermore, using phonological cues results in greater accuracies than using orthographic cues.

Next, we turn our attention to the performance of networks that always contained n-grams as cues and a combination between n-grams (orthographic and phonological bigrams and trigrams) and an abstract cue (**Figure 3**, below the line). In this way, we contrasted to what degree the inclusion of a specific abstract cue changed the F-scores of these networks.

We find that adding the number of syllables as cue does not provide the network with additional information when stress is counted from left ($\beta = 0.07196$, $sd = 0.21586$, $z = 0.333$, $p = 0.739$), but it improves the network when stress is counted from right ($\beta = 0.6132$, $sd = 0.1629$, $z = 3.763$, $p = 0.000168$). Providing the network with a more abstract representation of syllable number, namely parity, did not change network accuracies significantly ($\beta = 0.08391$, $sd = 0.22977$, $z = 0.365$, $p = 0.715$). The same holds true for higher level information about inflectional information ($\beta = 0.02881$, $sd = 0.23778$, $z = 0.121$, $p = 0.904$). However, in line with multiple psycholinguistic studies, we find that adding information about syllable structure significantly improves the network: this improvement is significant when stress position is counted from right ($\beta = 0.4256$, $sd = 0.1620$, $z = 2.628$, $p = 0.00859$), but failed to be significant, when it is counted from left ($\beta = 0.2074$, $sd = 0.2066$, $z = 1.004$, $p = 0.315$).

3.2 Cross-validation of results

It is very likely that the networks's high classification accuracy is only due to the fact that it was tested on seen data and might differ when tested on unseen data. We therefore performed twenty 10-fold cross-validation analyses, in which we trained the network on 90% of the word forms that were randomly selected and tested the network's performance on the remaining 10%.

For the sake of simplicity, we wanted to test networks with a specific combination of n-grams and abstract cues in the cross-validation. The question therefore arose what cues should be tested? As can be seen in **Figure 3**, the addition of number of syllables and syllable structure to n-grams as cues increased the F-scores of the networks. This indicates that the addition of these cues adds predictive information to the networks. Accordingly, we tested number of syllables and syllable structure, in addition to phonological or orthographic bigrams/trigrams as cues. As outcomes, we tested both stress assignment from left and stress from right.

Figure 4 illustrates how the weighted F-scores differed depending on the n-gram type (bigrams left, trigrams right, orthographic vs. phonological illustrated by line type) provided to the network (ignoring effects of number of syllables and syllable structure, as these were present in all networks), averaged across the 10 cross validation trials. As can be seen, trigrams as cues outperform bigrams and phonological cues outperform orthographic cues. The best network – stress counted from left discriminated by means of phonological trigrams as cues – yields an average F-score higher than 0.9. Accordingly, we are confident that the results reported in **Figure 3** are valid (given the clear differences between the types of n-grams, we did not perform any statistical analysis).
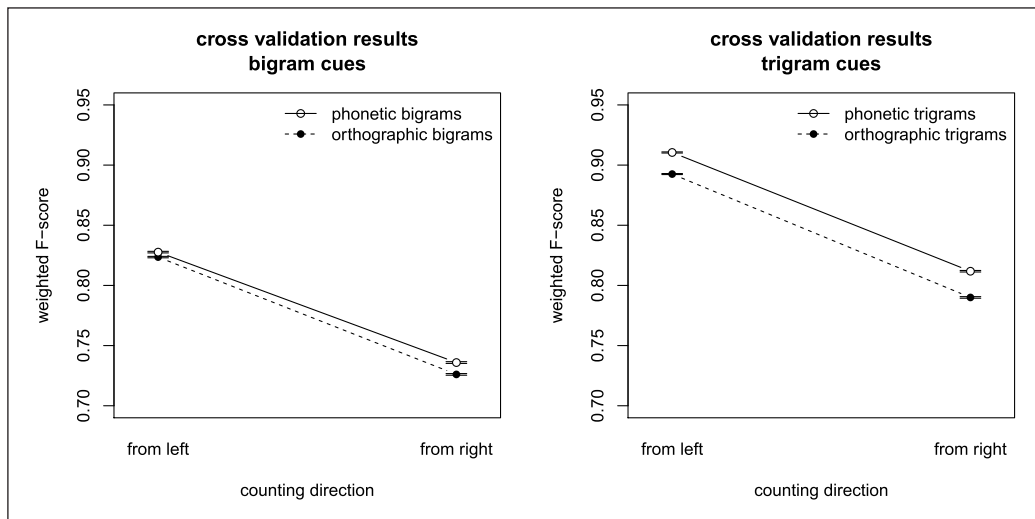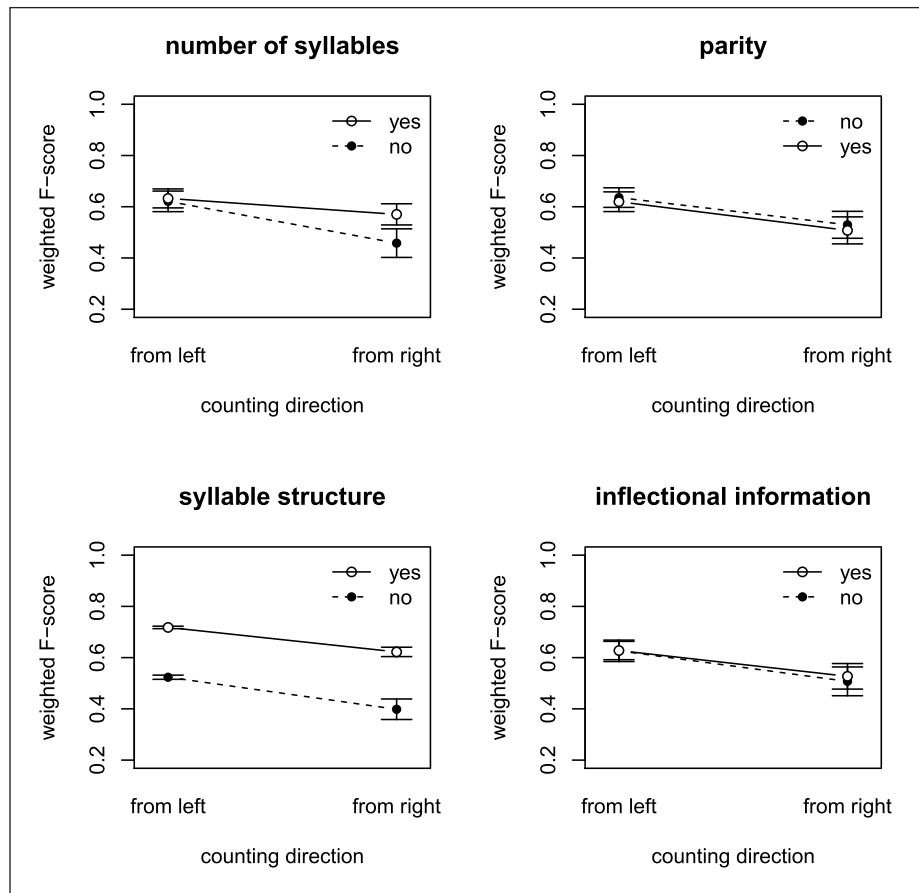

Figure 4: Average weighted F-scores depending on types of cues in twenty 10-fold cross validation trials (trained with 90%, tested with the remaining 10%).

Having tested the networks' classification performance when n-grams were used as cues during training, we focus in the next section only on the effects of abstract cues.

3.3 Effect of abstract cues

Most of the networks discussed above contained n-grams as cues. The question, however, arises to what degree abstract features are able to predict stress positions when the networks do not provide the fine-grained sublexical n-gram cues. The accuracies for these networks are illustrated in **Figure 5**. Overall, we see that the F-scores for networks missing n-grams as cues drop to around 0.6, i.e. a decrease of 0.1 to 0.3 in F-score in comparison to when n-grams are included. **Figure 5** shows that when stress is counted from left, the performance is equally good independently of whether the number of syllables is provided or not. However, when stress is counted from right, then F-scores strongly decrease without the number of syllables.

Figure 5: Average weighted F-scores from networks depending on types of cues for networks *without* n-grams as cues.

3.4 Directionality, morphological complexity, and stress position

Our analysis so far showed that stress counting from left outperformed counting from right. This, however, contrasts with assumptions from generative phonology and previous evidence from corpus analyses (e.g., Domahs et al. 2014b) and psycholinguistic experiments (Domahs et al. 2014a). However, note that these analyses focused on morphologically simple words consisting of two to four syllables. To investigate whether previous findings were due to the restricted set of stimuli used, we performed additional analyses on such a subset of trisyllabic, morphologically simple words.

When predicting stress in trisyllabic morphologically simple words, weighted F-scores range between 0.65 and 0.74 when stress is counted from right. However, when stress is counted from left, accuracy drops dramatically to an F-score between 0.34 and 0.36. This result supports the above mentioned conclusions from previous work that in trisyllabic, morphologically simple German words stress assignment seems to operate from right.

To investigate the influence of morphological complexity in more detail, we calculated the differences between the two directions of counting depending on morphological complexity, stress position and the number of syllables in a word. The differences per position and word length (number of syllables) are illustrated in **Figure 6** for words with 1 to 8 syllables (9 and 10 were excluded due to data sparsity). Positive differences indicate that the *count from left* network yielded a higher accuracy than the *count from right* network; negative differences indicate that the *count from right* network yielded higher accuracy than the *count from left* network.
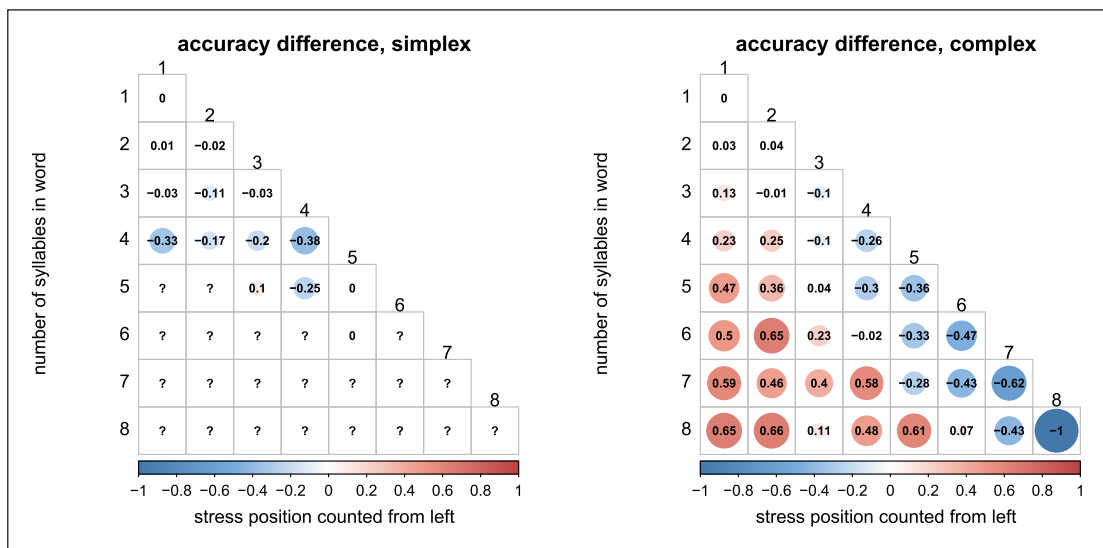
Figure 6: Differences in classification accuracy between the counting from left and counting from right models, depending on number of syllables per word (y-axis) and the stress position counted from left. Positive values indicate that the *count from left* network yielded a higher accuracy than the *count from right* network; negative values that the *count from right* network yielded higher accuracy than the *count from left* network. Positive values are marked red, negative blue. Question marks indicate missing information.

We first turn our attention to morphologically simple words. There is a negligible difference between *count from left* and *count from right* in disyllabic words. In trisyllabic words, the negative difference between 3%P and 11%P indicates that the *count from right* network performs better than the *count from left network*. This is also the case for words with four and five syllables, with very large negative differences emerging in four-syllable words. The negative differences indicate that *count from right* yields a higher classification accuracy than *count from left*. This result mirrors findings in the previously discussed literature on stress assignment.

The picture changes when we inspect morphologically complex words. Whether the *count from left* network or *count from right* network performs better depends on the distance of the stress position from the left and the right edge of the word. The *count from right* network performs better

when stress is located on the last or second last syllable of the word (this is true for most cases). The *count from left* network performs better when stress is located at the onset of words up to four syllables and in the centre of the longer words. This division of labour between the types of counting and stress position raises the question whether a network considering counting from left *and* counting from right jointly would better represent stress position. We tested this hypothesis by training a network to classify such a stress position on the basis of number of syllables and phonological bigrams and trigrams. However, this network yielded worse classification accuracies than the equivalent network trained to classify stress position *counted from left* (with accuracy being 2.8%P and 4.9%P worse, for bigrams and trigrams respectively). This network performed 7.8%P and 6.4%P better than the equivalent *counted from right* network. In conclusion, while the performance of networks considering different counting directions depends on the position of stress in the word, overall *counting from left* is the more adequate algorithm for a corpus in which the majority of words is morphologically complex (compare number of tokens in **Figure 6**, top).

3.5 Reflections of suffixation

Having discussed how the networks classify stress on the basis of cues, two further questions arise: First, on what basis are the networks able to accomplish this task? Second, how is information about relations between cues and outcomes represented in the network? We will answer these questions by illustrating how weight strength in the network represents the informativity of cues about a specific stress position. We first illustrate this for selected suffixes, following the example of (Arndt-Lappe et al. 2022). Subsequently, we consider weight distributions for all word final cues in the corpus. For the sake of simplicity, we will use orthographic cues for this illustration.

In German, a set of word final graphemes or grapheme sequences that act as suffixes almost never attract stress to the final syllable. These include <-n, -l, -r, -m, -t, -s> following a schwa (i.e. <e>), or unstressed syllables like <-te>, represented by the trigrams <en#, el#, er#, em#, et#, es#, te#>. Likewise, word final syllables that end in <-ee, -ie, -ll, -ion, -on>, represented by the trigram cues <ee#, ie#, ll#, on#>, attract word final stress. This rough classification is supported by the percentage of words that end in word final stress in the above listed sequences: <en#> 0.2%, <el#> 1.9%, <er#> 1.8%, <em#> 3.3%, <et#> 18.9%, <es#> 4.2%, <te#> 0.0%, <ee#> 53.6%, <ie#> 61.7%, <ll#> 37.2%, <on#> 81.2%. From this follows that trigram cues representing words that tend to divert stress from the final syllable should have a low or negative connection weight to final stress. By contrast, (pseudo-) suffix cues should have a high positive weight for final stress.

Figure 7 demonstrates that our assumption is correct. The schwa-syllable cues all have negative weights associated with the final syllable (stress from right = 1). Note that these cues have a positive weight associated to the penultimate syllable (stress from right = 2), indicating that they predict penultimate stress very well. Cues for <-ee, -ie, -ll, -on> endings all have

positive weights for the final syllable, supporting stress on the final syllable when these cues are present. By contrast, they have negative weights for the penultimate syllable, disfavoring this stress position.
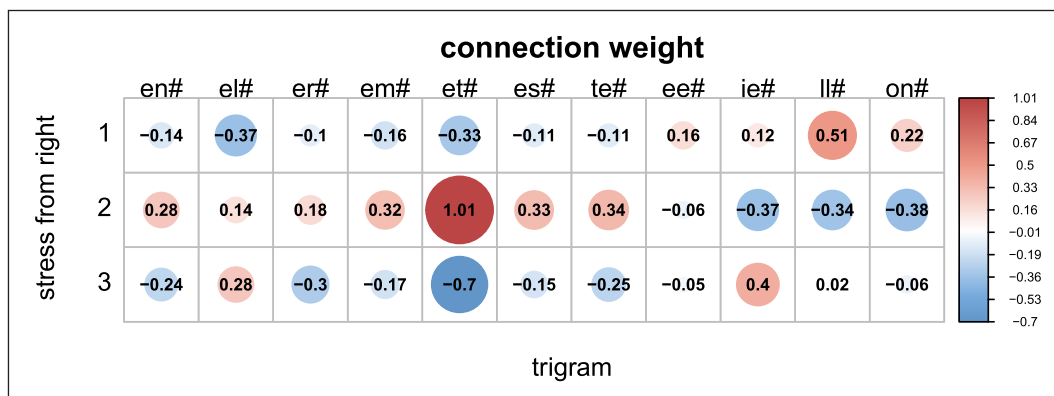

Figure 7: Connection strength between selected word final cues (x-axis) and stress position counted from right. Positive numbers + red circles indicate positive weights, supporting the selection of an outcome; negative numbers + blue circles indicate negative weights, supporting the non-selection of an outcome.

These examples demonstrate how the network represents systematic relations between cues and outcomes that are present in a language. However, for this example we selected cues for which we know that they systematically co-occur with final or penultimate stress, respectively. What about all other word-final cues in our corpus for which the relation is less clear?

To test this question, we inspected the connection weight of all word-final cues to all stress positions in the final, penultimate and antepenultimate syllables. We analyzed the relationship between a cue's probability of co-occurring with a specific stress position and the cue's connection weight to a specific stress position in the network. The results of this inspection are illustrated in **Figure 8**: (a) illustrates this relationship for antepenultimate stress; (b) and (c) illustrate this relation for penultimate and final stress, respectively. The positive regression line demonstrates that higher co-occurrence probability with a stress position yields a higher weight between a cue and an outcome. However, this is not consistently the case since weights are not equivalent to probability. The reason for this is because, as we have elaborated in the Introduction, weights represent how well a cue predicts an outcome and depend on both co-occurrence and non-occurrence between cues and outcomes. Due to this cue competition, some cues never co-occur with a stress position, as indicated by a probability of zero, yet the learning algorithm attributes them positive weights (this phenomenon is called *spurious excitement* (Kapatsinski 2021), which we will discuss in more detail in the Discussion section). Other cues always co-occur with a specific stress position, indicated by a probability of one. Most of these cues have positive

weights to that specific stress position. But note that the weights are distributed along the broad continuum with some weights falling even into the negative domain. This distribution of weights between positive and negative values for cues that occur with a specific outcome in 100% of the time illustrates a very important aspect of error-driven learning: neither individual cues nor cue combinations can predict an outcome with absolute certainty. We will discuss this relation between cues and stress position in more detail in the Discussion section, too.
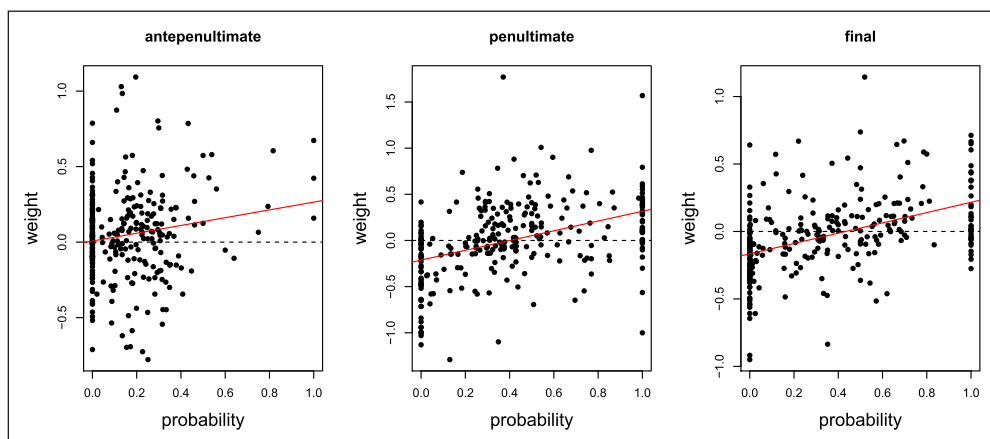
Figure 8: The x-axis illustrates the probability with which a word-final cue co-occurs with a specific stress position in a word (illustrated in columns). The y-axis illustrates the cue's connection weight to that specific stress position. See text for more details.

4 Discussion

4.1 Summary

Traditional linguistic theories typically predict stress position on the basis of discrete and abstract cues (e.g., Hayes 1995; Féry 1998; Kager 1999; Wiese 2000; Domahs et al. 2014b). In the present study, we contrasted these traditional approaches with a less abstract approach to stress assignment. We investigated how well a simple two-layered learning network (Naïve Discriminative Learner, NDL, Arppe et al. 2018) can predict stress position in German word forms when trained with an error-driven learning mechanism (Rescorla & Wagner 1972). We regard the prediction of the network to represent the assignment process.

We tested how different types of cues changed network performance. On the one hand, we tested abstract cues that are typically used in traditional approaches such as number of syllables, syllable structure, or morphological information. On the other hand, we tested cues that are more naïve about such abstract structures: n-gram cues representing a word's phonological or orthographic form (i.e. the word form). Overall, we found that assignment from right to left yields a higher accuracy in morphologically simple words than in morphologically complex

words while assignment from left to right yields better results for morphologically complex words. Moreover, we found that the naïve approach using word forms as cues to stress position outperforms a more abstract approach to stress assignment. In the following sections, we will discuss our results in more detail.

4.2 Counting direction

Previous analyses and experimental studies on stress assignment argue that word stress in German is assigned by anchoring the direction of assignment at the word offset – i.e. stress is assigned from right to left (Vennemann 1991; Giegerich 1985; Féry 1998; Jessen 1999; Domahs et al. 2014b). In parametric accounts following Hayes' principles of stress assignment (Hayes 1995), metrical trochaic feet are supposed to be constructed starting from the right edge of a word. Psycholinguistic evidence for German has been interpreted in favor of this assumption (Domahs et al. 2014a). However, in these accounts directionality is typically defined over morphologically simple words. When it comes to morphologically complex words, stress assignment is additionally guided by a number of morphological-phonological regularities that hold for different types of morphologically complex words or affixes, making stress assignment quite complex. Note, however, that NDL is agnostic with regard to both the prosodic and morphological units of complex words. Nevertheless, across the whole range of morphologically complex words, it was more successfull to assign stress when starting from left.

The fact that for morphologically simple words stress assignment was more successful when starting from right, while for complex words when starting from left also depending on word length implies that when a naïve approach to stress assignment is used, as in the present study, both directions of assigning stress are possible. These findings may also have cognitive implications. At an abstract level, certain phonological and orthographic characteristics of words such as word length and certain strings of graphemes and phonemes, may be associated with morphological structures, such as compounding or affixation. Linguistic theory suggests that different kinds of abstract morphological structure is in turn associated with specific stress positions (Giegerich 1985; Wiese 2000; Alber & Arndt-Lappe 2020). Our study, though, implies that stress position can be inferred directly from the formal characteristics of phoneme or grapheme sequences themselves, bypassing an abstract representation of morphological structures. In order to produce main stress in a specific position, this direct (non-abstract) relation has to be learned by speakers and readers during development (see e.g. Burani et al. 2014).

4.3 Naïve vs. abstract cues

We have found that NDL networks can predict stress position in German words based on their sequences of phonemes or graphemes with a very high accuracy (cf. **Figure 3**). This result

indicates that most cues about the stress position are provided by the word forms themselves, as Arndt-Lappe et al. (2022) have already demonstrated for English word stress. Moreover, this finding is compatible with the assumption that the position of word stress can be determined by means of analogy (Guion et al. 2003; Arciuli et al. 2010; Burani et al. 2014; Moore-Cantwell 2020).

How can these findings be related to previous studies that based stress assignment predominantly on abstract cues? When networks were trained with individual abstract cues and their combinations, prediction accuracy was about 30%P lower in contrast to when word form cues were provided. This means that, in spite of this drop, stress position could still be predicted fairly well with average accuracy of 60% when using abstract cues only. When we contrasted different types of cues, we found that the number of syllables and syllable structure, i.e. those types of cues typically used in traditional approaches, strongly improved network accuracy (cf. **Figure 5**). This means that, in line with traditional approaches, abstract cues are able to predict stress position to a certain degree, when this is the only type of cues the network has to associate with stress position.

Thus, our findings do not contradict abstract approaches to stress assignment per se (for German: Domahs et al. 2008; Janßen & Domahs 2008; Röttger et al. 2012; Domahs et al. 2014b; for Dutch: Kager 1989; Trommelen & Zonneveld 1999; Zonneveld & Nouveau 2004; for English: Liberman & Prince 1977; Giegerich 1985; Trommelen & Zonneveld 1999). They rather expand the types of cues that need to be taken into account when stress position is considered. Concretely, by using word forms as cues, we have demonstrated that the position of word stress can be determined on the basis of more fine grained pieces of cues than typically assumed. Our study furthermore demonstrates that once these more fine grained pieces of information are taken into account, word stress may not be 'calculated' or assigned through rules. Instead, it is more likely to be decided on the basis of the discriminative power of word form cues through an analogical process.

Another aspect concerns how abstract cues emerge from fine grained cues. Abstract cues are those pieces of information that are predictive about an outcome when variability across all instances of word forms is ignored. We argue that cues for stress position can be learned through prediction and prediction error (Ramscar et al. 2010, 2013b). Our networks demonstrate that it is possible to learn to associate those parts of the word form that predict a particular stress position and learn to ignore – i.e. unlearn – those parts of the word form that do not predict a particular stress position. Through this process, more abstract information such as syllable structure or number of syllables can be conceived as the results of generalizing information about a stress position across many, similar fine grained cues. (But note that we have not modelled this process since this would entail either a different type of cue-to-outcome set-up or a network with

hidden layers). These considerations raise the question of where exactly in the word form the decisive cues are located. We will discuss this question in the next section.

4.4 Individual vs. contextual cues

Traditional approaches to stress assignment argue that the position of main stress is decided on the basis of a set of very few cues. For example, when the final syllable is light, the penult is most likely stressed (Domahs et al. 2014b). From a probabilistic perspective, the assumption would be that the probability of assigning a stress position based on a particular cue is proportional to the frequency that this cue occurs with that stress position. However, from the learning perspective taken in the present study, this relation is not so straightforward. It turns out that how well a cue predicts a certain stress position (measured by its weight) is not directly proportional to the probability in which it co-occurs with that stress position (cf. **Figure 8**). This becomes visible when cues are considered that are unambiguous from a probabilistic perspective but obtain weights which are contrary to our naïve understanding how associations are formed. In the following, we will discuss how these cues are represented in the network by also discussing how the network learns specific associations.

On the one hand, there are some cues that, even though they never occur with a particular stress position, have positive weights. This positive weight (erroneously) indicates that a stress position should occur with that particular cue. The attribution of positive weights between cues and outcomes that never co-occur together is called *spurious excitement* (Kapatsinski 2021). This is a known property of the Rescorla-Wagner and Danks learning equations. It is considered to be the result of a particular mathematical constellation during the estimation of weights (exactly how they come into being is illustrated in Tomaschek 2020). Kapatsinski (2021) argues that spurious excitement is an error in the learning equations and that humans do not show this behavior.

On the other hand, there are some cues that have negative connection weights to that stress position, even though they occur 100% of the cases with a particular stress position. In such cases, the weight (again erroneously) indicates that this stress position should not occur with that cue. This effect is rooted in the weight estimation mechanism, too. Recall that how well a cue predicts an outcome depends on the relationship between how well the outcome matches the prediction on the basis of all the present cues. This means that the mechanism that estimates the weight takes into account how often a specific cue serves to predict other outcomes in addition to the amount of cues present during the learning event (which is argued to be cognitively plausible, see various publications by Ramscar and colleagues). From this follows that cues in isolation, as often analyzed in traditional approaches, can never predict a particular stress position with absolute certainty – even if they tend to occur frequently or even always with that position, as is the case for schwa syllables or certain suffixes (especially given that whether or not

a word final phoneme sequence constitutes a suffix is the result of top-down abstract analyses). Individual cues always have to be considered in the context of other cues. It could be that even though one particular cue predicts, for instance, final stress, other cues in that word rather predict penultimate stress. Thus, how well a cue predicts a particular outcome always has to be considered in relation to how strongly the other cues predict that particular outcome. In our modelling, we consider our cue-to-outcome structure to be a model of the cognitive mechanisms of stress assignment used by speakers. But what are the exact implications of the present results for psycholinguistic theories of stress assignment? This will be discussed in the next section.

4.5 Cognitive aspects

In our modelling, we have used the network in a *generative* fashion: This means that we provided it with a set of cues and asked it to predict the position of stress in a word. This network set-up implies that stress is assigned as part of a procedural speech production process, similar to psycholinguistic models of both speaking and reading that draw a distinction between the production of familiar words and unfamiliar or pseudo-words (e.g. Levelt et al. 1999; Caramazza et al. 2001; Grainger et al. 2012). In these models, unfamiliar words involve an online computation of the word form, including the assignment of its stress position. By contrast, familiar words involve access to word forms stored in the mental lexicon, potentially including its stress position.

The network set-up used in the present study seems to reflect the online computation of stress position based on an input of segmental strings (phonemes or graphemes). In fact, it learned stress assignment on a subset of words and was successfully able to generalise to previously unseen words which, in principle, could also be pseudo-words. Thus, these algorithms could be considered to represent sublexical processing. However, they do not support accounts of sublexical processing that claim any kind of default stress position (e.g. Eisenberg 1991; Levelt 1999; Wiese 2000). Clearly, we did not find evidence for the existence of a default stress position in German. Importantly, the two-layer representation of inputs and outputs does not reflect a notion of lexical processing in speech production or reading that involves static lexical representations in the mental lexicon. By contrast, they are compatible with the notion of lexical representations as states of a cognitive system that arise dynamically as a consequence of external or internal stimuli (Baayen et al. 2019). This means that the representation of a word's stress position in the mental lexicon is best considered as knowledge which is not only influenced by frequency of use but also by the position of stress in the ever changing constellation of neighboring words. As such, the presented set-up can be taken to model the abstraction or generalization process that is necessary to yield sublexical mappings from lexical input.

Acknowledgements

Competing interests

The authors have no competing interests to declare.

References

Alber, Birgit. 1998. Stress preservation in German loan words. *Phonology and Morphology of the Germanic Languages. Tübingen: Niemeyer* 113–114. DOI: https://doi.org/10.1515/9783110919769.113

Alber, Birgit. 2020. Word stress in Germanic. In *The Cambridge Handbook of Germanic Linguistics,* 73–96. DOI: https://doi.org/10.1017/9781108378291.005

Alber, Birgit & Arndt-Lappe, Sabine. 2020. Morphology and metrical structure. In *Oxford Research Encyclopedia of Linguistics*. DOI: https://doi.org/10.1093/acrefore/9780199384655.013.614

Arciuli, Joanne & Monaghan, Padraic & Seva, Nada. 2010. Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language* 63(2). 180–196. DOI: https://doi.org/10.1016/j.jml.2010.03.005

Arndt-Lappe, Sabine & Schrecklinger, Robin & Tomaschek, Fabian. 2022. Stratification without morphological strata, syllable counting without counts – modelling English stress assignment with Naive Discriminative Learning. *Morphology*. DOI: https://doi.org/10.1007/s11525-022-09399-9

Arnold, Denis & Tomaschek, Fabian & Sering, Konstantin & Lopez, Florence & Baayen, R. Harald. 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. DOI: https://doi.org/10.1371/journal.pone.0174623

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

Baayen, R. Harald & Chuang, Yu-Ying & Shafaei-Bajestan, Elnaz & Blevins, James 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 linear discriminative learning. *Complexity* 2019. Publisher: Hindawi. DOI: https://doi.org/10.1155/2019/4895891

Baayen, R. Harald & Milin, P. & Durdevic, Dusica & Hendrix, Petar & Marelli, Marco. 2011. An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. *Psychological Review* 118(3). 438–481. Publisher: American Psychological Association. DOI: https://doi.org/10.1037/a0023851

Baayen, R. Harald & Piepenbrock, Richard & van Rijn, Hedderik. 1993. *The CELEX Lexical Database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Bröker, Franziska & Ramscar, Michael. 2020. Representing absence of evidence: why algorithms and representations matter in models of language and cognition. *Language, Cognition and Neuroscience* 1–24. Publisher: Taylor & Francis. DOI: https://doi.org/10.1080/23273798.2020.1862257

Burani, Cristina & Paizi, Despina & Sulpizio, Simone. 2014. Stress assignment in reading Italian: Friendship outweighs dominance. *Memory & Cognition* 42(4). 662–675. DOI: https://doi.org/10.3758/s13421-013-0379-5

Caramazza, Alfonso & Costa, Albert & Miozzo, Michele & Bi, Yanchao. 2001. The Specific-Word Frequency Effect: Implications for the Representation of Homophones in Speech Production. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 27. 1430–1450. DOI: https://doi.org/10.1037/0278-7393.27.6.1430

Cribari-Neto, Francisco & Zeileis, Achim. 2010. Beta regression in R. *Journal of Statistical Software* 34. 1–24. DOI: https://doi.org/10.18637/jss.v034.i02

Daelemans, Walter & Gillis, Steven & Durieux, Gert. 1994. The acquisition of stress, a dataoriented approach. *Computational Linguistics* 20(3). 421–451.

Danks, David. 2003. Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology* 47. 109–121. DOI: https://doi.org/10.1016/S0022-2496(02)00016-0

Davis, Sally M. & Kelly, Michael H. 1997. Knowledge of the English noun–verb stress difference by native and nonnative speakers. *Journal of Memory and Language* 36(3). 445–460. Publisher: Elsevier. DOI: https://doi.org/10.1006/jmla.1996.2503

Domahs, Frank & Grande, Marion & Huber, Walter & Domahs, Ulrike. 2014a. The direction of word stress processing in German: evidence from a working memory paradigm. *Frontiers in Psychology* 5. 574. DOI: https://doi.org/10.3389/fpsyg.2014.00574

Domahs, Ulrike & Plag, Ingo & Carroll, Rebecca. 2014b. Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics* 17(1). 59–96. Publisher: Springer. DOI: https://doi.org/10.1007/s10828-014-9063-9

Domahs, Ulrike & Wiese, Richard & Bornkessel-Schlesewsky, Ina & Schlesewsky, Matthias. 2008. The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25(1). 1–36. DOI: https://doi.org/10.1017/S0952675708001383

Eisenberg, Peter. 1991. Syllabische Struktur und Wortakzent. Prinzipien der Prosodik deutscher Wörter. *Zeitschrift für Sprachwissenschaft* 10(1). 37–64. DOI: https://doi.org/10.1515/zfsw.1991.10.1.37

Eisenberg, Peter. 2016. *Grundriss der deutschen Grammatik: Band 2: Der Satz.* Springer-Verlag.

Ernestus, Mirjam & Neijt, Anneke. 2008. Word length and the location of primary word stress in Dutch, German, and English. *Linguistics* 46(3). 507–540. DOI: https://doi.org/10.1515/LING.2008.017

Féry, Caroline. 1998. German word stress in optimality theory. *Journal of Comparative Germanic Linguistics* 2(2). 101–142. DOI: https://doi.org/10.1023/A:1009883701003

Giegerich, Heinz J. 1985. *Metrical Phonology and Phonological Structure. German and English.* Cambridge: Cambridge University Press.

Grainger, Jonathan & Lété, Bernard & Bertand, Daisy & Dufau, Stéphane & Ziegler, Johannes C. 2012. Evidence for multiple routes in learning to read. *Cognition* 123(2). 280–292. Publisher: Elsevier. DOI: https://doi.org/10.1016/j.cognition.2012.01.003

Guion, Susan G. & Clark, J. J. & Harada, Tetsuo & Wayland, Ratree P. 2003. Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46(4). 403–426. DOI: https://doi.org/10.1177/00238309030460040301

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hoppe, Dorothée B. & Hendriks, Petra & Ramscar, Michael & van Rij, Jacolien. 2022. An exploration of error-driven learning in simple two-layer networks from a discriminative learning perspective. *Behavior Research Methods*. DOI: https://doi.org/10.3758/s13428-021-01711-5

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

Janssen, Ulrike. 2003. *Wortakzent im Deutschen und Niederländischen, Experimentelle Untersuchungen zum deutschen und niederländischen Wortakzent*. Düsseldorf: PhD dissertation, Heinirch-Heine-Universität Düsseldorf.

Jessen, Michael. 1999. German. In Van der Hulst, Harry (ed.), *Word Prosodic Systems in the Languages of Europe*, 515–545. Berlin: Mouton de Gruyter.

Jurafsky, Dan. 2000. *Speech & Language Processing*. India: Pearson Education.

Kager, René. 1989. *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: ICG Printing.

Kager, René. 1999. *Optimality Theory*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511812408

Kapatsinski, Vsevolod. 2021. Learning fast while avoiding spurious excitement and overcoming cue competition requires setting unachievable goals: reasons for using the logistic activation function in learning to predict categorical outcomes. *Language, Cognition and Neuroscience*. 1–22. Publisher: Routledge eprint. DOI: https://doi.org/10.1080/23273798.2021.1927120

Levelt, Willem J. M. 1999. Models of word production. *Trends in Cognitive Sciences* 3(6). 223–232. http://www.sciencedirect.com/science/article/pii/S1364661399013194. DOI: https://doi.org/10.1016/S1364-6613(99)01319-4

Levelt, Willem J. M. & Roelofs, Ardi & Meyer, Antje S. 1999. A theory of lexical access in speech production. *The Behavioral and Brain Sciences* 22(1). DOI: https://doi.org/10.1017/S0140525X99001776

Liberman, Mark & Prince, Alan. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.

Mattys, Sven & Samuel, Arthur G. 2000. Implications of stress-pattern differences in spoken word recognition. *Journal of Memory and Language* 42(4). 571–596. DOI: https://doi.org/10.1006/jmla.1999.2696

Mengel, Andreas. 2000. *Deutscher Wortakzent: Symbole, Signale*. Doctoral dissertation, TU Berlin.

Moore-Cantwell, Claire. 2020. Weight and final vowels in the english stress system. *Phonology* 37(4). 657–695. DOI: https://doi.org/10.1017/S0952675720000305

Nieder, Jessica & Tomaschek, Fabian & Cohrs, Enum & de Vijver, Ruben van. 2021. Modelling Maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience* 37(3). 381–402. Publisher: Routledge eprint. DOI: https://doi.org/10.1080/23273798.2021.1977835

Nieder, Jessica & van de Vijver, Ruben & Tomaschek, Fabian. 2022. "All mimsy were the borogoves" – a discriminative learning model of morphological knowledge in pseudo-word inflection. *Language, Cognition and Neuroscience*. 1–18. Publisher: Routledge eprint. DOI: https://doi.org/10.1080/23273798.2022.2127805

Nixon, Jessie S. 2020. Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition* 197. 104081. DOI: https://doi.org/10.1016/j.cognition.2019.104081

Nixon, Jessie S. & Tomaschek, Fabian. 2020. Learning from the Acoustic Signal: Error-Driven Learning of Low-Level Acoustics Discriminates Vowel and Consonant Pairs. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society,* vol. 42, 585–591. https://cognitivesciencesociety.org/cogsci20/papers/0105/index.html.

Nixon, Jessie S. & Tomaschek, Fabian. 2021. Prediction and error in early infant speech learning: A speech acquisition model. *Cognition* 212. 104697. DOI: https://doi.org/10.1016/j.cognition.2021.104697

Olejarczuk, Paul & Kapatsinski, Vsevolod & Baayen, R. Harald. 2018. Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard* 4(s2). 20170020. Publisher: De Gruyter. DOI: https://doi.org/10.1515/lingvan-2017-0020

Ramscar, Michael. 2021. A discriminative account of the learning, representation and processing of inflection systems. *Language, Cognition and Neuroscience*. 1–25. Publisher: Routledge eprint. DOI: https://doi.org/10.1080/23273798.2021.2014062

Ramscar, Michael & Dye, Melody & Klein, Joseph. 2013a. Children Value Informativity Over Logic in Word Learning. *Psychological Science* 24(6). 1017–1023. DOI: https://doi.org/10.1177/0956797612460691

Ramscar, Michael & Dye, Melody & McCauley, Stewart. 2013b. Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language* 89(4). 760–793. https://www.jstor.org/stable/24671957. DOI: https://doi.org/10.1353/lan.2013.0068

Ramscar, Michael & Yarlett, Daniel. 2007. Linguistic Self-Correction in the Absence of Feedback: A New Approach to the Logical Problem of Language Acquisition. Cognitive Science 31(6). 927–960. Publisher: Blackwell Publishing Ltd. DOI: https://doi.org/10.1080/03640210701703576

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

Rastle, Kathleen & Coltheart, Max. 2000. Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language* 42(3). 342–364. DOI: https://doi.org/10.1006/jmla.1999.2687

Rescorla, Robert & Wagner, Allan. 1972. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, Abraham H. & Prokasy, William Frederick (eds.), *Classical Conditioning II: Current Research and Theory*, 64–69. New York: Appleton Century Crofts.

Röttger, Timo B & Domahs, Ulrike & Grande, Marion & Domahs, Frank. 2012. Structural factors affecting the assignment of word stress in German. *Journal of Germanic Linguistics* 24(1). 53–94. DOI: https://doi.org/10.1017/S1470542711000262

Schiller, Niels O & Jansma, Bernadette M. & Peters, Judith & Levelt, Willem J. M. 2006. Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes* 21(1–3). 112–140. DOI: https://doi.org/10.1080/01690960400001861

Speyer, Augustin. 2009. On the change of word stress in the history of German. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 131(3). 413–441. DOI: https://doi.org/10.1515/bgsl.2009.051

Tomaschek, Fabian. 2020. The wizard and the computer: An introduction to preprocessing corpora using R. Tech. rep. PsyArXiv. https://psyarxiv.com/jsv38/. Type: article. DOI: https://doi.org/10.31234/osf.io/jsv38

Tomaschek, Fabian & Plag, Ingo & Ernestus, Mirjam & Baayen, R. Harald. 2019. 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. Publisher: Cambridge University Press. DOI: https://doi.org/10.1017/S0022226719000203

Tomaschek, Fabian & Ramscar, Michael. 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: https://doi.org/10.3389/fpsyg.2022.754395

Trommelen, Mieke & Zonneveld,Wim. 1999. Word stress in English and Dutch. In Van der Hulst, Harry (ed.), *Word Prosodic Systems in the Languages of Europe*, 132–133.

Turk, Alice E & Jusczyk, Peter W & Gerken, LouAnn. 1995. Do English-learning infants use syllable weight to determine stress? *Language and Speech* 38(2). 143–158. DOI: https://doi.org/10.1177/002383099503800202

Van Oostendorp, Marc. 2012. Quantity and the three-syllable window in Dutch word stress. *Language and Linguistics Compass* 6(6). 343–358. DOI: https://doi.org/10.1002/lnc3.339

Vennemann, Theo. 1991. Skizze der deutschen Wortprosodie. *Zeitschrift für Sprachwissenschaft* 10(1). 86–111. DOI: https://doi.org/10.1515/zfsw.1991.10.1.86

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

Zonneveld, Wim & Nouveau, Dominique. 2004. Child word stress competence: an experimental approach. In Kager, René & Pater, Joe & Zonneveld, Wim (eds.), *Constraints in Phonological Acquisition,* 369–408. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486418.012