Stratification effects without morphological strata, syllable counting effects without counts – modelling English stress assignment with Naive Discriminative Learning

Sabine Arndt-Lappe[1] · **Robin Schrecklinger[1]** · **Fabian Tomaschek[2]**

Abstract

Stress position in English words is well-known to correlate with both their morphological properties and their phonological organisation in terms of non-segmental, prosodic categories like syllable structure. While two generalisations capturing this correlation, directionality and stratification, are well established, the exact nature of the interaction of phonological and morphological factors in English stress assignment is a much debated issue in the literature. The present study investigates if and how directionality and stratification effects in English can be learned by means of Naive Discriminative Learning, a computational model that is trained using error-driven learning and that does not make any a-priori assumptions about the higher-level phonological organisation and morphological structure of words. Based on a series of simulation studies we show that neither directionality nor stratification need to be stipulated as a-priori properties of words or constraints in the lexicon. Stress can be learned solely on the basis of very flat word representations. Morphological stratification emerges as an effect of the model learning that informativity with regard to stress position is unevenly distributed across all trigrams constituting a word. Morphological affix classes like stress-preserving and stress-shifting affixes are, hence, not predefined classes but sets of trigrams that have similar informativity values with regard to stress position. Directionality, by contrast, emerges as spurious in our simulations; no syllable counting or recourse to abstract prosodic representations seems to be necessary to learn stress position in English.

✉ S. Arndt-Lappe
arndtlappe@uni-trier.de

R. Schrecklinger
robin.schrecklinger@uni-trier.de

F. Tomaschek
fabian.tomaschek@uni-tuebingen.de

[1] Universität Trier, Universitätsring 15, 54296 Trier, Germany

[2] Seminar für Sprachwissenschaft, Universität Tübingen, Wilhelmstraße 19-23, 72074 Tübingen, Germany

1 Introduction

Stress position in English words is well-known to correlate with both their phonological and morphological properties. For example, stress is often penultimate in morphologically simplex nouns with a heavy penultimate syllable, as illustrated by the word 'agenda' *a.gén.da*. In derived words with a so-called stress-preserving suffix, stress is always on the same syllable as it is in the base word. For example, *éffort-less* is stressed on the same syllable as *éffort*, in spite of the fact that the word *effort-less* has a heavy penultimate syllable. By contrast, stress in derived words with so-called stress-shifting suffixes may be on a different syllable than it is in the base word (e.g. *adópt – adopt-ée*).[1] In classic rule-based or constraint-based accounts within (morpho-)phonological theory, the correlations just mentioned are usually interpreted in terms of a causal relation within the grammatical architecture. Thus, stress algorithms are assumed to be sensitive to the weight of the penultimate syllable, and to the morphological status of suffixes. This presupposes that also lexical representations must explicitly code relevant information that the algorithm can refer to (or, in an output-oriented framework like Optimality Theory, a function *Gen* that generates a candidate set containing this information). For the examples just mentioned, this means that lexical representations must encode:

– syllabification and weight information, generalising over different sounds by syllabic position;
– the morphological status of suffixes as stress-preserving or stress shifting.

With regard to both types of information, there is a rich literature about how these properties are best formally represented (cf. e.g. Newell, 2021 for a recent summary and proposal). The complexity and degree of abstraction of pertinent formalisations has also provided a major problem for alternative, usage-based theories of linguistic generalisation (Bybee, 2011, 2001, 2002), whose architectures do not refer to abstract properties so much.

In the present paper we ask whether the encoding of abstract phonological and morphological properties in lexical representations does indeed form the prerequisite for any theory of how stress is assigned in English words. We will approach our question by conducting a series of simulation studies with a computational model. To reduce the complexity of the question, the focus of our study will be confined to two descriptive generalisations that play a prominent role in virtually all formal accounts.

The first is the principle of directionality (Hayes, 1982; see Pater, 2000 for an optimality-theoretic account for English; see Kager, 2012 for a typological overview and a discussion of different modelling options within Optimality Theory and

---

[1]The distinction between 'stress preserving' and 'stress shifting' affixes (Fudge, 1984) is largely co-referent with other dichotomies, such as that of 'cohering' – 'non-cohering', 'class I' – 'class II', 'level I' – 'level II', 'stem-level' – 'word-level' (Booij, 1983; Siegel, 1974; Booij & Rubach, 1987; Kiparsky, 1982a, 2015; Bermúdez-Otero, 2018).

Table 1 Examples of stress assignment in long English words. Main stress is indicated by an acute accent, syllable boundaries are marked by '.'

a.	Ka.la.ma.zóo
b.	Mo.non.ga.hé.la
c.	Ha.ma.me.li.dán.the.mum

Alber, 2020 for an overview on Germanic languages). Phonological generalisations about stress position are usually thought to be directional in the sense that they count syllables from a word edge. This also means that they crucially rely on word representations that incorporate syllables as abstract units of prosodic organisation. The relevant word edge in English is usually assumed to be a three-syllable window at the right word edge.

The examples in Table 1 illustrate the principle. Main stress is indicated by an acute accent, syllable boundaries are marked by '.'. A description of stress patterns that is in line with the principle of right directionality will refer to stress as being on the word-final syllable in (1a), on the penultimate syllable in (1b), and on the antepenultimate syllable in (1c). This generalisation captures the fact that, with the exception of compound words and many words bearing stress-preserving suffixes, main stress in English words always lands on one of the last three syllables of the word. However, the idea that main stress assignment is directional or even mono-directional in nature is not without problems.

For example, Hammond (1999, 318ff) and McCully (2003) argue that both right alignment and left alignment play a role for main stress assignment in English. Furthermore, attempts to empirically verify edge alignment face the problem that the number of English morphologically simplex words that are longer than three syllables is rather low. Another problem is that different affixes require different adjustments to the syllable counting generalisation about stress position. This phenomenon, among others, has led scholars to assume that the English lexicon is stratified, with the different strata representing different morphological categories. A final problem with directionality is that stress generalisations based on syllable counting from a word edge are not without exceptions. In English, for example, words like *állegory* and *árbitrary*, which do not contain any stress-preserving suffixes, are exceptions to the generalisation that main stress is assigned to one of the three syllables at the right word edge (e.g. Burzio, 1994; Trevian, 2007).

Morphological stratification in this sense is the second generalisation that we will be concerned with in this article. Specifically, we will focus on suffixal strata, often referred to as 'stress-preserving' and 'stress-shifting'. Stress-shifting suffixes fall into two different subgroups: those which themselves attract stress (often called 'auto-stressed') and those which do not; most of the latter suffixes are 'pre-stressing', which means that main stress is on the syllable immediately preceding the suffix. Table 2 provides examples of all three classes. The suffixes *-ness* and *-ly* are examples of stress-preserving suffixes (2a, but cf. Trevian, 2007 for some cases of stress shift with *-ly*), *-ity* and *-ical* are pre-stressing stress-shifting suffixes (2b), and *-ee* and *-ese* are auto-stressed stress-shifting suffixes.

The existence of stress-preserving and stress-shifting suffixes has prominently been used as evidence in favour of stratal approaches to morphology-phonology interaction such as Lexical Phonology and Morphology (Kiparsky, 1982b) and Stratal

Table 2 Examples of
a) stress-preserving,
b) stress-shifting and
c) auto-stressed suffixation. The
accent indicates the stressed
syllable

	Derivative	Base
a.	háppiness	háppy
	políteness	políte
	símilarly	símilar
	prodúctively	prodúctive
	extrémely	extréme
b.	similárity	símilar
	productívity	prodúctive
	verbósity	verbóse
	metaphórical	métaphor
	symmétrical	sýmmetry
c.	interviewée	ínterview
	adoptée	adópt
	Portuguése	Pórtugal
	Japanése	Japán

Phonology (Bermúdez-Otero & McMahon, 2006; Bermúdez-Otero, 2012, 2018). These approaches assume that English morphology is organised into two (or more) different strata, with interleaving phonological and morphological modules. The difference between stress preserving and stress shifting suffixes is then modelled in terms of the point in time when a suffix is attached to its base word or stem. So-called stress-shifting suffixes are attached before phonological stress rules have applied, stress-preserving suffixes are attached after stress rule application. Other approaches model the stress behaviour of different types of affixes in terms of affix-specific rule or constraint systems (cf. e.g. Stanton & Steriade, 2014). However, the exact nature of the interaction of phonological and morphological factors in a stratified lexicon is a much debated issue in the literature. Empirically, it is well-known that existing proposals (stratal or non-stratal) integrating phonological and morphological factors fall short of convincingly predicting stress position when tested on data sets of words, both actual and nonce words. Furthermore, attempts to quantify accuracy of predictions are comparatively rare, often limited to subsets of the lexicon and operationalised in such way that they cannot be compared. In the following, we will discuss these studies.

Zamma (2012) developed a model that includes variable constraint ranking. Accuracy is measured in terms of the number of predicted rankings that conform to attested words (cf. Zamma, 2012, chpt. 6 for discussion). Domahs et al. (2014) provide a statistical analysis of the predictive power of syllable structural factors in morphologically simplex words, both nonce words and existing words. Simplex words are also studied by Moore-Cantwell (2016, chpt. 4); the study investigates the match between a constraint-based MaxEnt model (Goldwater & Johnson, 2003) that includes lexically specific constraints, and lexical distributions. Dabouis et al. (2017) investigate the predictive power of both phonological and morphological factors for stress in some 5,000 verbs extracted from Jones (2006)'s English Pronouncing Dictionary.

All works cited show that the phonological and morphological factors they use in the analysis can explain a large portion of the data, but also admit to considerable leakage. In all pertinent accounts, it is thus assumed that stress assignment is subject to lexical idiosyncrasy to some extent (cf. Alber, 2020 for a recent overview of the literature on stress in Germanic languages, including English). It is also unclear how these studies can be compared, since all of them use different kinds of baselines, constraints and evaluation metrics.

Another open question concerns how language users become aware of these principles. One suggestion is that learning of stress takes into account abstract representations of the prosodic and morphological structure of words, and on the basis of constraints that operate on the basis of those abstract representations (cf. e.g. Moore-Cantwell, 2016 for a recent OT model; cf. Pearl et al., 2016 for a comparison of the learnability of classic pertinent approaches). Abstract representations include syllables, morae, metrical feet, and the morphological stratum affiliation (e.g. level 1 or level 2) of affixes. Constraints include constraints on edge alignment, on the relation between syllable weight and stress, on extrametricality, as well as on the stressability of affixes. While some researchers suggest that these representations and constraints are innate, there is evidence that they themselves are the result of learning.

For example, Jarmulowicz and colleagues (Jarmulowicz, 2006; Jarmulowicz et al., 2008) found in elicited production experiments that school children aged 7–9 years reached surprisingly low accuracy rates for derived words with the pre-stressing suffixes *-ity* and *-ic* (between 24% for nonce words and 70% for high-frequency words), but not for derived words with the pre-stressing suffix *-ion* (between 74% for nonce words and 93% for high-frequency words). Adult control groups performed at ceiling level in the same experiments (cf. Jarmulowicz et al., 2008, 222, Table 2). These findings demonstrate that, despite the uniformity of stress among words with pre-stressing suffixes and the hypotheticised categorical nature of phonological constraints often presupposed in the literature, the learning task is far from trivial. Moreover, this finding also suggests that at least some morphological stress alternations are acquired rather late, with learner productions heavily correlated with frequency and, hence, storage in the Mental Lexicon. It is at present unclear how especially the latter can be accounted for under the assumption that learning of stress takes place on the basis of abstract phonological and morphological principles, which should be in place, once acquired.

In the present paper, we will test if an alternative answer is viable, which is in line with usage-based theories of linguistic generalisation (Bybee, 2011, 2001, 2002). By 'usage-based' approaches we mean a group of theories that share the assumption that properties such as stress are associated with and may even emerge from the distributional characteristics of words and sub-word units in the Mental Lexicon. For stress assignment, this means that language users store words that they encounter with their stress pattern, and assign stress to words they have not encountered before on the basis of the distribution of stress patterns among stored words.

So far, only few attempts have been made to test this idea on stress assignment data with the help of computational implementations of usage-based models. One notable exception is Daelemans et al. 1994's study of stress position in Dutch monomorphemes. A key challenge for a usage-based model of stress assignment is the definition and selection of input features provided to the computational model. Unlike

formal phonological theory, computational modelling approaches usually rely on flat and non-nested structures. This does not seem compatible with generalisations about stress assignment, which, as we have seen above, rely on abstract and elaborate representations of phonological and morphological structure. Daelemans et al. 1994 used an Instance-Based Learning algorithm (Aha et al., 1991) to successfully predict stress in a large database of Dutch monomorphemic words. Comparing different encodings of input features at different levels of phonological abstraction, they find that the model performed best when trained on actual phonemic representations, as compared to a more abstract coding of syllable weight or of syllabic constituents (rhymes) deemed relevant in the formal literature. Still, their implementation shares some basic assumptions with the formal literature. Thus, directionality was implemented in the model by aligning words in the input coding at their right word edge and by defining stress categories in terms of the position of the stressed syllable from the right word edge (i.e. the final, penultimate, or antepenultimate syllable). Also, morphological structure was eliminated as a potential confound as complex words were excluded from their dataset.

The present paper will take a more radical approach, presenting a computational model with the plain form of words without any explicit, higher-level information about potential positions of stress or morphological structure. Our interest is to see if and how directionality and morphological stratification in English can be learned by such a model. The particular implementation that we will use is Naive Discriminative Learning ('NDL', Arppe et al., 2018, cf. Baayen et al., 2011 for the theoretical underpinnings). The advantage of using this model is that it is a relatively simple algorithm (a two-layer network), whose output can be analysed in a way that is meaningful for linguistic research. Another advantage is that NDL is trained using a cognitively plausible learning rule ('Discriminative Learning', cf. below for details). Based on a series of simulation studies we will show that neither directionality nor stratification need to be assumed to be a-priori properties of words or constraints in the lexicon. Stress can be learned solely on the basis of very flat word representations in terms of trigrams, by a system that is not given any explicit information about directionality or the morphological class affiliation of constituent affixes. Instead, morphological stratification emerges as an effect of the model learning that informativity with regard to stress position is unevenly distributed across all trigrams constituting a word. Morphological affix classes like stress-preserving and stress-shifting affixes are, hence, not predefined classes but sets segmental strings that have similar informativity values with regard to stress position. Directionality, by contrast, emerges as spurious in our simulations; no syllable counting or recourse to abstract prosodic representations seems to be necessary to learn stress position in English.

The paper is structured as follows. We will first introduce our computational framework in Sect. 2. Sect. 3 will explain the methodology of our simulation experiments. The simulations will then be discussed in Sect. 4, in two steps. We will first be concerned with directionality (Sect. 4.1), and then with morphological strata (Sect. 4.2). In each section, we will present both general simulation outcomes and an in-depth analysis of our experiments, which shows why the algorithm makes the predictions it does. The paper ends with a summary and conclusion in Sect. 5, which will also discuss the implications for linguistic theory.

2 Discriminative learning and the error-driven learning rule

Different approaches to training a neural network are available. Due to hidden layers or complex learning algorithms, as is the case in deep neural networks and recurrent neural networks (Graves & Schmidhuber, 2005; Graves et al., 2013), such trained networks are usually hard to interpret from a cognitive perspective. We therefore used a two-layer neural network that is trained with a simple error-driven learning rule (Rescorla & Wagner, 1972; Rescorla, 1988; Ng & Jordan, 2002; Widrow & Hoff, 1960), implemented in Naive Discriminative Learning (the package 'NDL' as implemented in R, Arppe et al., 2018).

The error-driven learning rule mathematically formalizes general cognitive mechanisms assumed by the cognitive theory of *Discriminative Learning* (Ramscar & Yarlett, 2007; Ramscar et al., 2010, 2013a). According to this theory, learners build cognitive representations of their environment by establishing associations between events in their environment on the basis of prediction and prediction error. The algorithm formalizes this by establishing *association weights* between input features (henceforth *cues*) and classes or categories (henceforth *outcomes*) that co-occur in events. To name an example, in English the word final letter sequence '-ize' serves as a cue to the outcome 'verb', and the word final letter sequence '-ical' serves as a cue to the outcome 'adjective'.

According to Discriminative Learning theory, learning is shaped by prediction and prediction error. Learning leads to an increase in association weights between a cue and an outcome every time that the predicted outcome co-occurs with a cue (such as '-ize' in the word 'conceptualize', which is a verb). By contrast, error is negative and decreases association weights whenever the predicted outcome does not occur (such as 'ize' in the noun 'size'). As a result, weights and associations (and the resulting representations) are constantly updated on the basis of new experiences. The strength of the adjustment depends a) on the number of cues that are present in a learning event and b) on the size of the error between the prediction emerging from the cues and the actual outcome in the learning event. This gives rise to *cue competition*, during which cues compete for being informative about an outcome. As a result of learning through continuous prediction and error, cognitive representations emerge. An in-depth description of the theory can be found in (Ramscar et al., 2013a; Linke & Ramscar, 2020; Ramscar, 2021); a description of the NDL model can be found in Baayen et al. (2011); an overview of how different cue-to-outcome structures affect learning can be found in Hoppe et al. (2022).[2]

The error-driven learning rule has been shown to successfully model and predict a number of important effects observed in animal learning (Rescorla, 1988) and human learning. In human experiments, this has been accomplished by training simple two-layer neural networks using the error-driven learning rule. Subsequently, network measures such as activations were computed the basis of the networks and used as

---

[2]This formalization of learning differs from other theories of learning, such as Bayesian models (Kleinschmidt & Jaeger, 2015), or distributional learning models (Wanrooij et al., 2014, 2015; Werker et al., 2012; Terry et al., 2015). The latter class of models assumes that learners learn the frequency of occurrence of co-occurrences and the resulting distributions. For a review of the differences between distributional learning and error-driven learning in different linguistic contexts see Kapatsinski (2018).

predictors for human behavior. For example, Ramscar et al. (2010) demonstrated how the presentation order of cues and predicted events during learning affects the strength of learning. Learning is more effective when, for example in 'wug' experiments, the orthographic (or acoustic) word precedes the corresponding picture than when the picture precedes the word. This effect was also reported for phonetic learning (Nixon, 2020) and inflectional learning (Hoppe et al., 2020). Nixon (2020) demonstrated that a new cue for an outcome is blocked from learning, once another cue has already been learned as informative about an outcome. This finding mirrors the 'blocking effect' in animal learning studies first demonstrated by Kamin (1968).

In addition, the error-driven learning rule successfully models aspects of child language acquisition (Ramscar et al., 2010, 2011, 2013b,a), acquisition and usage of allomorphic suffixes (Divjak et al., 2021), reaction times in lexical decision tasks (Baayen et al., 2011; Milin et al., 2017b), self-paced reading (Milin et al., 2017a), phonetic characteristics depending on their morphological function (Tomaschek et al., 2019; Saito et al., 2020; Tomaschek & Ramscar, 2022; Schmitz et al., 2021), auditory comprehension (Baayen et al., 2016; Arnold et al., 2017) and acoustic single-word recognition (Shafaei-Bajestan & Baayen, 2018). Furthermore, it was applied in modelling early phonetic learning (Nixon & Tomaschek, 2021, 2020) and morphological processes of pluralization in Maltese (Nieder et al., 2022a,b).

To summarize, the association weight between a cue and an outcome is formed through the experience with other cues and outcomes that have been encountered during the learning history in both production and comprehension. The weight represents the support which a specific cue can provide for a specific outcome. Cognitive representations of grammatical structures emerge from the association weights between every encountered cue and every encountered outcome. In this model principles like the principle of directionality and stratification have no independent status as constraints on representations or grammatical outputs. The question then arises if and how the model can emulate and explain the empirical effects that have traditionally been ascribed to these mechanisms.

3 Methods

For our simulation experiments, we trained NDL to discriminate stress positions and then used the trained network to predict stress positions. The material for the simulations ($N = 33,407$ lemmas, i.e. word types[3]) was obtained from the CELEX lexical database of English (Baayen et al., 1993). This data set served as both the training set and the test set. We performed our analysis in two steps. In a first step we focused on directionality and investigated which cue structure best predicts the attested stress patterns. In a second step, we focused on morphological stratification and studied if and how exactly morphological strata emerged in our model. All data and scripts underlying the analysis presented in this paper are available online, at https://osf.io/8nbyj/. In what follows we discuss the methodological details of our modelling approach.

---

[3]Lemmas were extracted from the epl.cd file in CELEX.

The input cues on which we trained our models were bigrams and trigrams, derived from the orthographic representations of all lemmas in CELEX. Using orthographic representations as input thus means that our simulations may be conceived of as reading experiments in which readers are pronouncing English words, except that we abstract away from different word forms. Still, the choice of orthographic bigram and trigram cues clearly deserves discussion, as orthographic bigrams and trigrams per se are not cognitively plausible form representations of words in the Mental Lexicon. We use them here as proxies for word representations that make the fewest possible a priori assumptions about abstract representational units.

Bigram and trigram representations encode orthotactic information, i.e. sequential information about adjacent letters in words, without including formally defined syllables or syllable positions (cf. e.g. Baayen et al., 2011, 2016; Milin et al., 2017b; Tomaschek et al., 2019, for modelling approaches following a similar rationale). We tested which kind of cue structure best predicted stress position: letter bigrams (BG), or trigrams (TG), or both together (BGTG).[4]

Using orthographic rather than phonetic transcriptions allowed us to avoid the problem that in many English words, knowing the vowel quality is already predictive of stress, and at the same time to avoid making any potentially controversial assumptions about underlying phonological vowel qualities. This is because only a very restricted set of vowels can occur in surface realisations of unstressed English syllables, a phenomenon that is usually accounted for in terms of phonological processes ('vowel reduction'). The most common reduced vowel, schwa, is even restricted to exclusively occurring in unstressed syllables. For example, a common pronunciation of the word 'America' is [əmɛrəkə]. Given this sound structure, it is clear that the stress can only be on [ɛ], the only full vowel. Given the orthographic string <America>, however, all syllables are potentially stress-bearing. Providing the computational model with orthographic cues rather than with actual pronunciations, thus, serves to make its task more difficult.

The main conceptual problem with using orthographic cue representations, however, is that it seems to run counter the well-established idea that the assignment of stress is based on phonological, not orthographic representations. Apart from its methodological motivation, the cue structure we use in our experiment also provides an interesting test case for evidence documented in the literature that, contrary to what might be expected, phonological computation is not independent of orthography in English (cf. Montgomery, 2001, 2005; Giegerich, 1992, 1999; on English word stress cf. esp. Guierre, 1979, et seq., Arciuli & Cupples, 2006; Arciuli et al., 2010; Dabouis et al., 2018; Dabouis, 2022; Deschamps, 1994). Also the high error rates that Jarmulowicz and colleagues found in 7–9 year-old children's productions of complex words indicate that acquisition of morphological constraints on word stress (specifically: stress shift) is not completed before literacy (Jarmulowicz et al., 2008; Jarmulowicz, 2006, discussed in Sect. 1 above). We take this to mean that an influ-

---

[4]Theoretically, we could also use higher-order *n*-grams such as 4-grams or lower-order monographs. However, the longer the *n*-gram, the stronger the model is faced with a one-to-one mapping between cues and outcomes, which results in smaller cue-competition during training. By contrast, letter monographs miss out on orthotactic information.

ence of orthographic representations on the acquisition of stress patterns cannot be precluded.

We conclude that the jury is still out on what role exactly orthographic representations play as a cue to stress position in English. The fact that, as we will see below, orthographic cues are indeed very good predictors of stress position in English lends additional support to this being a topic worth pursuing in future research.

Stress position was coded as outcomes in our simulations. We implemented three different types of outcome structures. The first is a representation of the traditional account that the stress position in a word is counted from the offset of the word (henceforth *count from right* (e.g. Hayes, 1982; Pater, 2000; Alber, 2020, as discussed in Sect. 1 above). In order to examine the validity of this claim, we also tested two other ways of representing stress as outcomes in our model. The first is to count the stress position from the onset of the word (henceforth *count from left*). The second is to select the vowel letter present in the stressed syllable (henceforth *vowel*). The value of *count from right* varied between one and seven. *count from left* contained six values, ranging between stress on syllable number one and stress on syllable number six. *vowel* did not differentiate in which syllable the vowel was located and contained 59 different values in total. The high number of different values results from the fact that English orthography encodes vowels also by means of digraphs or trigraphs which are taken into account here.[5] Stress representations that involved syllable counting (*count from right* and *count from left*) were based on the phonetic transcription and syllabification provided in CELEX.

Take the word 'reluctantly' as an example. Its letter bigram cues are `#r`, `re`, `el`, `lu`, `uc`, `ct`, `ta`, `an`, `nt`, `tl`, `ly`, `y#`, its letter trigrams are `#re`, `rel`, `elu`, `luc`, `uct`, `cta`, `tan`, `ant`, `ntl`, `tly`, `ly#` (where # represents the word boundary). Crucially, the model is unaware what phone sequences the letter *n*-grams represent. The outcomes of the models, – called 'outputs' in the terminology of neural networks – represent the position of the stress. For 'reluctantly', this means that *count from left* is: 2; *count from right* is: 3, and *vowel* is: 'u'.

We compared nine different networks in terms of how well they predict stress in our data set. Each network was trained on a different combination of cue and outcome structures (3×3, i.e. bigram cues, trigram cues, and a combination of bigram and trigram cues with the outcome *count from left*, the outcome *count from right*, and the outcome *vowel*). We use the Danks Equilibrium Equations (Danks, 2003) to train the model, as implemented in the NDL package. After training, the network is evaluated in terms of whether it is able to discriminate among the outcomes on the basis of presented cues (typically from a word of interest). Thus, it is presented by a set of cues, e.g. `#re`, `rel`, `elu`, `luc`, `uct`, `cta`, `tan`, `ant`, `ntl`, `tly`, `ly#`, and has to select which of the potential outcomes (e.g. for *count from right* 1, 2, 3, 4, 5, 6, or 7) is best predicted by the cue set. This is achieved by means of an activation vector, summing up the association weights between the presented cues for each of the possible outcomes in the network. The outcome with the highest activation is the winner of the classification, thus the predicted stress position.

---

[5]Note that a small number of these vowel graphemes may in fact be heterosyllabic in some words.

One aspect in which the classification procedure just described seems to depart from a cognitively plausible procedure is that the set of possible outcomes is defined on the basis of the whole data set, not on the basis of a given input. This is different e.g. from formal approaches such as Optimality Theory (e.g. Pater, 2000; Zamma, 2012; Moore-Cantwell, 2016), in which the selected outcome is one of the candidate outputs given to the procedure for a specific input. The problem can be best exemplified with the help of monosyllabic words. Naively, it should not be too hard to find the stressed position in monosyllabic words. Whereas this line of thought is of course plausible in the real world, it is not necessarily in our model. This is because this kind of reasoning follows the misconception that the model takes into account the number of syllables in the cue set that is presented during the classification procedure when computing a predicted outcome. This is not the case in the simulations presented in this paper, at least not in a direct way. It is therefore even possible that due to cue competition and due to the distribution of weights, the network predicts a stress position which is incompatible with the true number of syllables in the presented word. For example, it is possible that the network erroneously predicts stress on the penultima for a monosyllabic word. The task that the model faces to find the right stress position is therefore more difficult in a way than what is plausible from a cognitive perspective. As we will see in Sect. 4.2, however, the number of cases in which NDL actually predicts stress to land on a non-existent syllable is very low in practice.

4 Results

4.1 Classification accuracy by cue-outcome structure

Each of our nine networks (cf. Sect. 3 above) was set the task of predicting stress position in all words from CELEX. As can be seen in Table 3, classification accuracies for all cue-and-outcome combinations range between 59.0% and 84.9%, i.e. highly above chance. As is clear from the table, the use of letter bigrams consistently yields a lower classification accuracy than the use of letter trigrams. Also, a combination of bigrams and trigrams did not improve classification accuracy. This means that letter trigrams are sufficiently informative about stress positions.[6]

Given that letter trigrams yield better classification accuracy, we focus on this cue structure in all subsequent analyses. We now inspect how it was used by the network to classify stress positions given different assumptions about directionality. Table 3 demonstrates that stress can be learned without syllable counting. The model trained

---

[6]The question arises why trigrams yield a higher accuracy than bigrams. Given that trigrams capture a larger portion of a word than bigrams, the uncertainty about the relation between cues and the stress position should be lower for trigrams than for bigrams. We assessed this uncertainty by calculating the entropy (Shannon, 1948) for each bigram and for each trigram in relation to the stress position. To obtain the entropy for each n-gram cue, we assessed how often each n-gram occurs with each stress position. To calculate entropy, we calculated the co-occurrence probability by dividing a cue's co-occurrence frequency with a particular stress position by the summed frequencies of that cue and all stress positions. This then is a measure of how informative a cue is w.r.t. stress position. We found that the average entropy in relation to stress position is significantly lower for trigrams ($H = 0.84$) than for bigrams ($H = 1.48$, $\delta H = 0.64$, $t = 24.54$, $df = 844.83$, p-value < 0.001).

Table 3 Percentage of correctly categorised stress positions in whole data set

Cue structure	Left	Right	Vowel
letter bigrams	71.4	59.0	72.1
letter trigrams	80.7	74.9	84.9
both together	80.6	74.9	84.8

to predict stress in terms of the orthographic vowel has the highest classification accuracy, followed by the model that was trained to predict stress from the left word edge. The weakest model is the one that was trained to predict stress from the right word edge. All differences between model accuracies are significant (*count from left* vs. *count from right*: $\chi^2 = 246.4$, $df = 1$, $p < 0.001$; *vowel* vs. *count from left*: $\chi^2 = 961.5$, $df = 1$, $p < 0.001$). Using trigrams as cues, we tested the *count from left* and *vowel* models in twenty cross-validation runs. In each run, we trained the models on 70% of the data that were randomly selected and tested on the remaining unseen 30% of the data. The average classification accuracy was 71.6% (sd = 0.005) in the *count from left* model and 75.8% (sd = 0.003) in the *vowel* model. Thus, even if the model has not encountered a word, it was able to predict its stress position with a fairly high accuracy.

Since the model had no a-priori information about morphological structure, and since suffixing influences stress position in English, it is not surprising that the *count from right* model showed only weak performance. This is because the descriptive generalisation that English stress always lands in a three-syllable window at the right edge is not true for complex words with so-called stress-preserving suffixes (cf. Sect. 1 above for discussion).[7] What is surprising, however, is that stress is best predicted by the *vowel* model, as none of the existing theories predicted this finding.[8]

Looking only at prediction accuracies, however, does not tell us much about why the models performed as well as they did. With regard to the vowel model, a very

---

[7]Readers might wonder why the *stress from left* model performed so well, given that prefixes should also have an influence on stress assignment, similar to that of suffixes. One might expect prefixes to be a problem because most English prefixes are 'stress-preserving' in the sense that, like stress-preserving suffixes, they neither attract main stress onto the prefix, nor do they cause main stress in the complex word to land on some syllable other than the syllable that is main-stressed in the base word. The fact that the presence of prefixes does obviously not lead to weaker classification accuracy in our models may be due to the fact that when a word is prefixed with one prefix it typically reoccurs with other, prosodically similar prefixes. For example, the adjective *interpretable* occurs with *un-*, *re-*, and *mis-*. In all cases, the prefixed word has the third stress position, which provides the model with strong support for this third position. Analogous to what we show for words with stress-preserving suffixes in Sect. 4.4, this support is most likely based on high activations of the bigram or trigram cues that represent the word stem rather than the individual prefix. In our example, this third position is the 'stress-preserving' position. A systematic analysis of stress predictions for prefixed words, however, is beyond the scope of this paper.

[8]As an anonymous reviewer pointed out, it might be likely that the higher classification accuracy in the *vowel* was due to repeated vocalic letters (e.g. <e> in 'elephant') and double-vocalic graphemes (e.g. <ea> in 'read' or <oo> in 'tool'). We tested this hypothesis by running additional control models in which we excluded words with a) double-vocalic graphemes indicating a stressed position, b) repeated vowels and c) a and b simultaneously from the training set. In all three cases, the models performed systematically better than our initial model, reflected by an increase of 2% to 8% of model accuracy (depending on the cues). Thus, double-vocalic graphemes and repeated vowels have a negative effect on classification accuracy.

likely confounding factor is that orthographic vowels may occur multiple times in words. It is thus unclear whether the high classification accuracy of the *vowel* model results from the fact that vowel repetition increases the probability of finding the correct stress position. In the following section, we turn to a more detailed statistical analysis of our three models with aim in mind to learn more about the potential confounding factors mentioned.

4.2 Word length and repeated vowels

In this section, we want to inspect how classification accuracy varies with structural and morphological information. To do so, we will focus on the models trained with trigram cues. One interesting question is how the models predict stress in words of different length. This is a crucial issue for our assessment of the directionality prediction. The average accuracy scores presented in the previous section suggest that stress position may indeed be learned without resorting to directionality. However, the distribution of different word lengths across the lexicon may constitute an important confound here, in two ways.

One is that the vast majority of English words are short, with monosyllables having a particularly large share in the vocabulary. Model accuracy on short words will therefore also have a large share in the general accuracy score. Recall that NDL does not have explicit information about word length in our dataset and, hence, it is, in principle, possible that a monosyllabic word is predicted to be stressed in other positions than the first. However, word length is correlated with the number of cues which are going to compete for influence in stress classification. As a consequence, there will be less competition between cues in short words than in long words. Especially monosyllables will also profit from the additional advantage that most of the time their cues will contain the vowel that is stressed, and there will be no competition from other cues containing vowels. We thus expect that classification accuracy will be very high for short words and will decrease in words with a greater number of syllables.

Another potential confound especially concerns the *vowel* model: Since the outcomes in this model do not differentiate in which syllable a vowel is located, a model prediction was counted as 'correct' when the prediction matched the vowel that is observed to be stressed in the word, independent of where in the word the vowel is located. As a consequence, the probability that the model correctly predicts stress is higher, when a word contains repetitions of the same vowel in different places (captured by *repeated vowels* in what follows). Accordingly, we expect classification accuracy to be higher when the word contains multiple instances of the same orthographic vowel rather than different vowels. Repeated vowels may, however, also affect the other models, *count from left* and *count from right*. This is because repetition of vowels will also reduce cue competition in these models.

Figure 1 illustrates classification accuracy of the three different models (*y*-axis) across the number of syllables in a word (*x*-axis) in interaction with *repeated vowels*. Accuracies were obtained by computing the appropriate contingency tables. The three models are represented by different line types, with the solid lines representing the *vowel* model. Black lines represent accuracies for the subset of words containing

Fig. 1 Model accuracies as
function of number of syllables
in word. Model types are
encoded by line type

number of syllables * repeated vowels
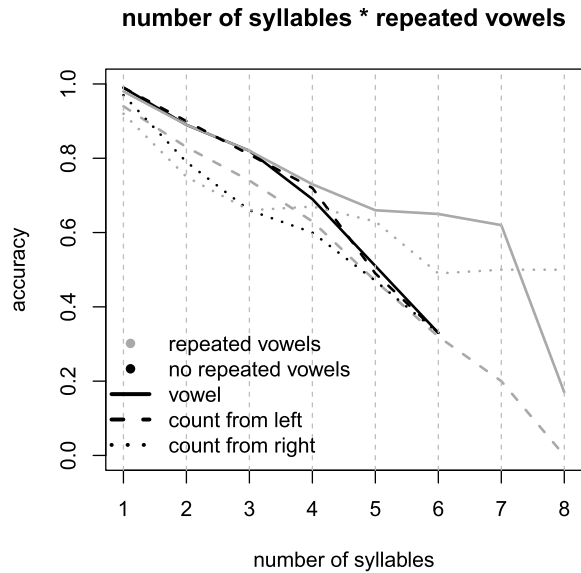
no repeated vowels, grey lines represent accuracies for the subset of words in which identical vowels were repeated. We computed accuracy scores for all three models (*vowel*, *count from left*, *count from right*) to allow readers to compare their performance on the exact same dataset.

We observe that model accuracies decrease as the number of syllables increases. Accuracy for monosyllabic words is highest (well above 95% in all subsets and models), which shows that, as predicted, NDL does hardly ever assign stress to nonexistent vowels or syllables in such words. Furthermore, even for four-syllable words, model accuracy stays above 75% for the two top models (*vowel*, *count from left*), which shows that, overall, prediction accuracy is highly satisfactory and clearly above chance level. The relatively low accuracies for very long words might strike readers versed in formal phonological theories as disappointing, especially because this seems to indicate that the model does not generalise to longer and rare words. However, we have to take into account in our interpretation of model accuracies that the model performs best for those types of words which it has seen most frequently. This type of behaviour has been repeatedly shown to reflect human learning behaviour in learning experiments (Ramscar et al., 2013a,b, 2010). One reason why accuracy drops so strongly for longer words therefore clearly lies in the way words of different length are distributed in our data set. The vast majority (roughly 95%) of all words in our data set are not longer than four syllables. While there are roughly 12,000 disyllabic (37%) and 8,850 trisyllabic words (27%), there are only 312 words with six syllables (1%), 40 words with seven syllables (0.12%), and only 6 words with eight syllables (0.02%). The distribution in our data set faithfully reflects the well-known distributional fact that the English vocabulary has a rather low proportion of long words. The model therefore did clearly learn assignment of word stress in this data set.

As we can also learn from Fig. 1, the hypothesis that high overall classification accuracy of the *vowel* model might be due to repeated vowels in the dataset is only partially correct. For words without repeated vowels the *count from left* model (dashed black line) performs equally well as the *vowel* model. By contrast, classification accuracies become worse when repeated vowels are present in the word (dashed grey line). Is this also the case in the *count from right* model? The answer to this question is: no. We observe that classification accuracies in the *count from right* model are on average worse across all word lengths than the other two models, even when words contain no repeated vowels.

Regarding the confounding effect of repeated vowels in the *vowel* model, for short words (two or three syllables) there is no difference in accuracy between words with or without repeated vowels. Accuracy is slightly lower in four-syllable words when no vowels are repeated, and strongly decreases in longer words in comparison to when a repeated vowel is present. Thus, as expected, there is an advantage for the classification when words contain repeated vowels in longer words. Note, however, that the relatively strong decrease in accuracy for five- and six-syllable words without repeated vowels is based on a very slim data set (86 five-syllable words, 3 six-syllable words). Virtually all words in this subset for which NDL provides an incorrect classification (42 words) are suffixed. By far the most frequent suffix is adverbial *-ly*, suffixed to adjectives ending e.g. in *-ical* (8x, e.g. *methodically*), *-able* (4x, e.g. *imperturbably*), *-ory* (2x, e.g. *perfunctorily*); other common suffixes in this set include *-ity* (6x, e.g. *jocularity*), *-ary* (5x, e.g. *rudimentary*), and *-ism* (4x, e.g. *obscurantism*). This is particularly interesting because long complex words ending in some of these suffixes (*-ly, -ary, -ism*) are well-known to vary between stress-shifting and stress-preserving behaviour in English (cf. e.g. Trevian, 2007, Bauer et al., 2013, chpts. 14, 15). Also, many erroneous NDL predictions do systematically pattern in ways that are not entirely implausible. For example, NDL predicts *méthodically, impérturbably, pérfunctorily, jócularity, rúdimentary*, and *obscúrantism*. Predicted stress in *obscúrantism* is actually a stress that is attested in English, but not in our database (Wells, 2008, s.v.; similarly, NDL predicts attested *círculatory* where our database has *circulátory*). Another relatively frequent type of error in this set is that stress preservation is predicted with stress shifting suffixes (e.g. predicted *jócularity, méthodical(ly)*, stressed on the same vowel as the embedded words *joke*, *méthod*); this occurs in 14 cases, in two of which the stresses predicted by NDL are listed as actual pronunciation variants in the *Longman Pronunciation Dictionary* (Wells, 2008).

4.3 Learning morphological stratification

4.3.1 Model accuracies and morphological structure

In the upcoming analysis, we want to gain an understanding of how morphological effects on stress assignment are represented by the model. The reasoning here is that the model's uncertainty about the relation between cues and outcomes should systematically vary with morphological complexity, and that this will be reflected in the model's classification accuracy of the stress position. To test this, we extracted the information included in CELEX about whether words are derived or simplex since

Fig. 2 Model accuracies as function of morphological complexity. Model types are encoded by point type
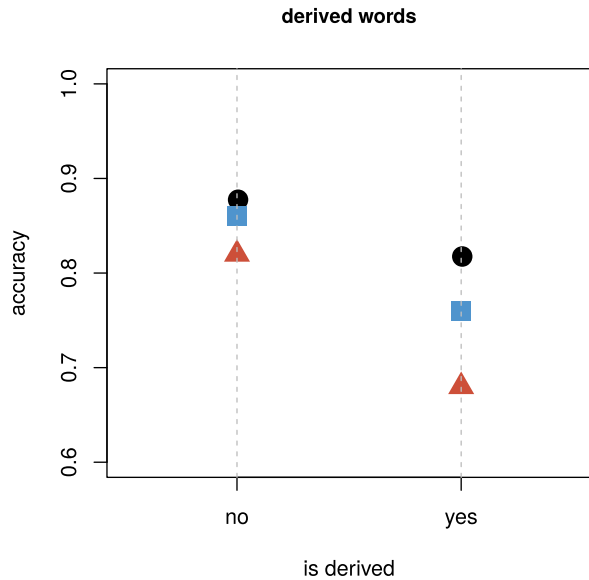
derived words

derivational processes can be stress-preserving or stress-shifting. This means that the variability of the stress position in derived words is very high, which should create more uncertainty about the stress position for the learning model. Accordingly, we expect accuracies for derived words to be lower than for underived words. Figure 2 supports our hypothesis in all three models.

Having demonstrated that our models classify derived and underived words with different degrees of accuracy, we will now investigate how accuracies vary with morphological type among derived words, i.e. stress preserving and stress shifting suffixation. For example, suffixes such as *-ion*, *-ity* or *-ical* (and their equivalent derivations) attract stress to the syllable preceding the suffix (*pre-stressing*), whereas suffixes such as *-ese*, *-teen* or *-ee* carry stress themselves (*auto-stressed*). Suffixes such as *-ness* or *-less*, by contrast, preserve basal stress (*stress preserving*, cf. Sect. 1 above for discussion).

We expect classification accuracy to be associated with the type of morphological structure. In words with stress preserving suffixes, the stress position should be strongly supported by the cues in the base. By contrast, in words with pre-stressing and auto-stressed suffixes, the stress position is different in derivatives and corresponding bases, which should result in more uncertainty about the stress position, due to competition between stress-supporting cues from the base and from the suffix. In words with stress-shifting suffixes, cues from the suffix support the stress position in the derived word, while the cues in the base will have to support multiple stress positions (at least one for the derived word and one for the base word). This is why we expect higher classification accuracy for stress preserving suffixes than for pre-stressing and auto-stressed suffixes. We do not make any predictions about the difference between auto-stressed and pre-stressing suffixes, as they both introduce greater cue competition than stress-preserving suffixes do.

Fig. 3 Model accuracies for derived words depending on morphological stress type. Model types are encoded by point type
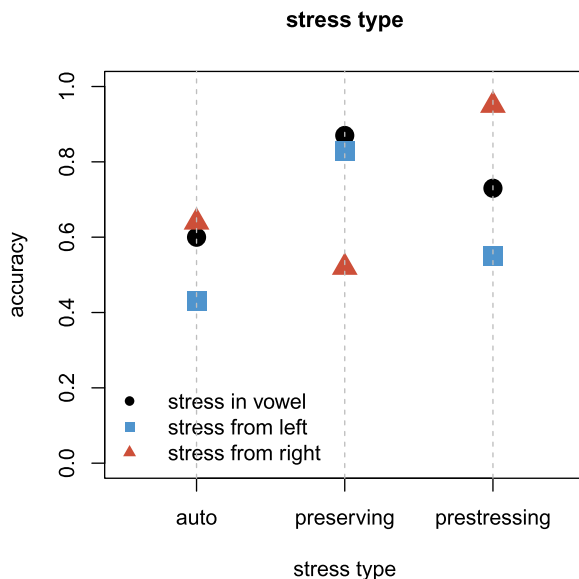
We tested these hypotheses with the help of a subset of 4,626 words that contained only words with clearly *stress preserving*, *pre-stressing*, and *auto-stressed* suffixes.[9] The stress preserving suffixes that we considered were *-ness* (as in *happi-ness*), *-less* (as in *piti-less*), and *-ly* (as in *happi-ly*). The pre-stressing suffixes that we considered were *-ion* (as in *constrict-ion* or *informat-ion*), *-ity* (as in *divin-ity*), and *-ical* (as in *satir-ical*). The auto-stressed group comprised the largest number of different suffixes, as these suffixes occur in much fewer different words in English than the suffixes belonging to the other two groups. By including a larger number of different suffixes in this group we made sure that we would have a sufficient number of data for analysis. These suffixes are *-ese*, *-teen*, *-ee*, *-ana*, *-esque*, and *-ette* (as in e.g., *Japan-ese*, *seven-teen*, *adopt-ee*, *Smithsoni-ana*, *Kafka-esque*). Suffixed words were extracted from our CELEX dataset on the basis of their orthographic forms (e.g. lemmas ending in the string 'ness', 'ity', 'ee'), but excluding lemmas that are not marked as morphologically derived in CELEX (e.g. excluding the lemma *city* from the set of *-ity* derivatives).

The accuracies in Fig. 3 support our hypothesis for the *vowel* and *count from left* models. In these models, stress preserving suffixes indeed yielded higher accuracies than auto stressing and prestressing suffixes. Note that prestressing suffixes yielded a higher classification accuracy than auto-stressed suffixes. A likely reason for this difference is that auto-stressed suffixes cause a stress shift in all pertinent words, and cues from the base have to compete for this stress position with the cues from the suffix. By contrast, stress in words with prestressing suffixes may often be in the same position as in their bases, which is why their stress position is stronger supported by

---

[9]Running the analysis on the whole set of derived words in our CELEX dataset rather than on a set of selected derivational categories was not an option. The reason is that assignment to stratal categories is not straightforward for all suffixes. Cf. e.g. Bauer et al. (2013, chpt. 9) for discussion.

the cues in the base. As an example, compare derivatives with pre-stressing *-ity*, some of which are stressed on the same syllable as their bases (e.g. *obésity*, *obése*) whereas others are not (e.g. *productívity*, *prodúctive*).

Next, we turn our attention to accuracies in the *count from right* model, for which the effect is reversed: preserving suffixes actually yield the lowest accuracy while auto-stressed and prestressing suffixes yield higher accuracies. This finding does not come as a surprise. Derivation by means of suffixation changes the number of syllables in a word. The syllable position remains the same when counted from left, but changes when counted from right. Accordingly, preserving suffixes create higher uncertainty in the *count from right* model than in the other models, reducing its accuracy. Note, however, that the *count from right* model yields a very high accuracy for prestressing suffixes, as they provide a highly informative cue about the stress position.

4.3.2 Strata and activation profiles

So far, we have shown that the NDL network is capable of learning stress position and that morphological structure is reflected in classification accuracies. We have argued that this is due to differences in the uncertainty between cues and outcomes. In the following section, we turn our attention to how these differences in uncertainty are reflected in the model. Specifically, we will inspect how the *vowel* model represents morphological stratification. We hypothesize that the network has indeed learned stratification of suffixes. Specifically, we assume that stratification will be mirrored in differences in the activation of the stress position coming from the stem and coming from the suffix.

The cues in suffixes that attract stress (*auto-stressed* suffixes) and suffixes which attract stress to the preceding syllable (*pre-stressing* suffixes) systematically indicate the stress position; the cues in the stem of such derivatives, by contrast, discriminate variable stress positions (i.e. those of the base word and those of its derivatives). Accordingly, suffix cues in stress-shifting derivatives should be better cues for the stress position than stem cues. From this we predict that, in NDL terms, stress-shifting suffixes will have a relatively higher activation than their stems. By contrast, the reverse should hold for stress-preserving suffixes. Here suffix cues should be weaker cues for stress position than stem cues. These suffixes should thus yield a lower activation than the stem.

We operationalized the relative support of stem and suffixes for the stress position by calculating the ratio between the activation of the suffix and activation of the stem for a word's stress position. Prior to the calculation, we rectified activations of stems by setting all activations smaller than zero to zero. The data showed a strong non-normal distribution with a peak at the centre and long left and right tails. This is why we used a non-parametric regression technique, quantile generalized additive models (Fasiolo et al., 2021), which allows to fit data without any assumptions about the distribution of the residuals.[10]

---

[10]Quantile generalized additive models allow to fit data from different quantiles of the dependent variable's distribution. In our analysis, we fitted the 50th percentile, i.e. the median.

Table 4 Summary for model fitting the activation ratio (suffix/stem) as a function of stress position shifts depending on suffixation. Intercept represents 'preserving' stress

	Estimate	std. error	z-value	p-value
(Intercept)	0.01	0.00	6.13	< 0.0001
morphological class = auto	0.42	0.04	11.41	< 0.0001
morphological class = prestressing	0.11	0.00	27.91	< 0.0001

Fig. 4 Average activation ratio between suffixes and stems depending on stress shifts due to suffixation

stress type effect on activation ratio
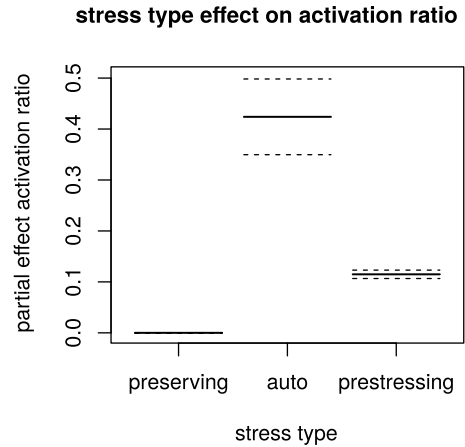
We used *activation ratio* as the dependent variable, and *stress position* as a factorial predictor (with *preserving* as the reference level). Table 4 reports the model summary. Figure 4 visualizes the results.

The intercept of the model, i.e. the average *activation ratio* for *stress preserving* suffixes, is 0.01. We see that the levels *auto stressed* and *prestressing* yield significantly higher activation ratios than the level *stress preserving*. However, average activation ratios are always below 1, which means that the stem is more strongly activated for the word's stress position than the suffix, regardless of its stratal affiliation. A very likely explanation is that, on average, stems have more cues ($\mu = 7.2$, sd = 2.5) than suffixes ($\mu = 2.9$, sd = 1.0). As a consequence, they contribute more weights for summation than suffixes, yielding overall higher activation scores.

In spite of suffixes having smaller activations than stems, the direction of the observed effects supports our hypothesis. Stratification is indeed mirrored in the activation profiles of derived words. *Auto stressed* and *prestressing* suffixes yield significantly higher activation ratios than *stress preserving* suffixes. In other words, stratification is reflected in the model in terms of systematic differences in the activation profiles of complex words.

5 Discussion

In the present study we set the Naive Discriminative Learner model (NDL, Baayen et al., 2011; Arppe et al., 2018) to the task of classifying stress position in simplex and

morphologically complex English words from the CELEX Lexical Database (Baayen et al., 1993). The representation of words that the model was given as input comprised bigrams and trigrams, i.e. flat representations that encode sequences of sounds or letters and as such, intrinsically encode phonotactic information. The most important lesson to be learned from our modelling experiments is that stress position in English words can be learned extremely successfully without assuming an a-priori setting of a directionality parameter, and without an a-priori specification of morphological strata in the Mental Lexicon.

With regard to directionality, we saw that orthographic vowels provide a better outcome structure for stress position than outcome structures based on syllable count from either word edge. This finding provides a substantial challenge to existing formal accounts, which all assume that directionality is an indispensable parameter in stress assignment.

The present findings also raise interesting questions about the role of orthography in stress assignment. In the present paper, orthographic representations were used as input simply because this offered a pragmatic solution to the problem that stress position and vowel quality are strongly correlated in English. Our simulations do, however, converge with previous work done in research on the acquisition of reading skills, which has provided support for the idea that orthography is indeed predictive of stress position (Arciuli et al., 2010; Abasq et al., 2019). While this was not our aim, the findings of the present study suggest that it might be so on a larger scale than expected. English orthography has already been shown to provide informative cues about morphological structure in words (Berg, 2013). Our study indicates that its graphemic structure discriminates stress position. In order to explore this issue further, however, more research is needed to better understand how exactly trigrams encode information that is relevant for language processing. Similar findings have already been reported before. For example, a comparison of studies employing NDL to model language processing tasks seems to suggest that trigrams are more informative cues than bigrams in some modelling tasks, but less successful in others (Baayen et al., 2011; Baayen & Smolka, 2020; Tucker et al., 2019). Why this is so, is not fully understood.

With regard to morphological stratification, we saw that differences between morphological categories can be understood as differences in the activation profiles of pertinent words. Activation profiles refer to the way in which the distribution of stored weights are skewed within a word, as a result of linguistic experience when learning complex words with their stress patterns. According to this account, what speakers learn when they learn words with stress-preserving suffixes is that cues for stress are relatively stronger in the base than in the suffix. Conversely, learning stress shift in this account means learning that cues for stress position are relatively stronger in the suffix. The model therefore offers an articulate hypothesis about what underlies stratification effects. This hypothesis is testable. One prediction worth exploring is that, if stratum-specific stress behavior is emergent from activation profiles, the model should predict stress variation to occur exactly in cases in which both the suffix and its stem are strongly activated (cf. Bell, 2015 for evidence that variation in English compound stress arises in similar situations). This prediction could be tested with the help of actual pronunciations of complex words. Another prediction is about the

acquisition of stress-shifting morphological categories. Thus, if it is true that learning stress shift involves learning higher activation of suffix cues, the simulations we presented here predict that, overall, stress shift should be acquired later than stress preservation. More importantly even, learning paths should be correlated with frequencies of pertinent morphological categories and with measures of how base and suffix cues compete for stress positions in words instantiating such categories. Both predictions could be tested by means of cross-sectional production studies as described by Jarmulowicz et al. (Jarmulowicz et al., 2008; Jarmulowicz, 2006), whose findings seem to already support the idea that stress shift is acquired relatively late, and that its acquisition is a gradual process. We leave testing these predictions (and others) to future research.

Data Availability https://osf.io/8nbyj/.

Code Availability https://osf.io/8nbyj/.

Declarations

Competing Interests n/a.

References

Abasq, V., Dabouis, Q., Fournier, J.-M., & Girard, I. (2019). The core of the English lexicon: stress and graphophonology. *Anglophonia 27*. https://journals.openedition.org/anglophonia/2317.

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1), 37–66.

Alber, B. (2020). Word-stress in Germanic. In M. T. Putnam & B. M. Page (Eds.), *The Cambridge handbook of Germanic linguistics*, Cambridge: Cambridge University Press.

Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, *59*, 920–948.

Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: corpus, behavioral, and computational investigations. *Journal of Memory and Language*, *63*, 180–196.

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., & Baayen, R. H. (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.

Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2018). ndl: naive discriminative learning. https://CRAN.R-project.org/package=ndl

Baayen, R. H., Milin, P., Durdevic, 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. Publisher: American Psychological Association.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database (CD-ROM). In *Linguistic data consortium*. Philadelphia: University of Pennsylvania.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: a proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, *31*(1), 106–128.

Baayen, R. H., & Smolka, E. (2020). Modeling morphological priming in German with naive discriminative learning. *Frontiers in Communication*, *5*, 17. Publisher: Frontiers.

Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: OUP.

Bell, M. (2015). Inter-speaker variation in compound prominence. *Lingue E Linguaggio*, *14*(1), 61–78.

Berg, K. (2013). Graphemic alternations in English as a reflex of morphological structure. *Morphology*, *23*(4), 387–408. Publisher: Springer.

Bermúdez-Otero, R. (2012). The architecture of grammar and the division of labour in exponence. In J. Trommer (Ed.), *The phonology and morphology of exponence – the state of the art* (pp. 8–83). Oxford: OUP.

Bermúdez-Otero, R. (2018). Stratal phonology. In S. J. Hannahs & A. R. K. Bosch (Eds.), *The Routledge handbook of phonological theory* (pp. 100–134). Abingdon: Routledge.

Bermúdez-Otero, R., & McMahon, A. M. (2006). English phonology and morphology. In B. Aarts & A. M. McMahon (Eds.), *The handbook of English linguistics* (pp. 382–410). Oxford: Blackwell.

Booij, G. E. (1983). Principles and parameters in prosodic phonology. *Linguistics*, *21*(1), 249–280.

Booij, G., & Rubach, J. (1987). Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry*, *18*(1), 1–44.

Burzio, L. (1994). *Principles of English stress*. Cambridge: Cambridge University Press.

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

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290. Publisher: Cambridge University Press.

Bybee, J. (2011). *Frequency of use and the organization of language*. New York: Oxford University Press.

Dabouis, Q. (2022). *English phonology and the literate speaker: some implications for lexical stress*. In N. Ballier, J.-M. Fournier, A. Przewozny, & E. Yamada (Eds.) *English word stress: theories, data and variation*. Oxford: OUP. To appear.

Dabouis, Q., Enguehard, G., Fournier, J.-M., & Lampitelli, N. (2018). *The English 'Arab rule' without metrics. Paper presented at the 26th Manchester Phonology Meeting, 24.–26.05.2018*, University of Manchester.

Dabouis, Q., Fournier, J.-M., Girard, I., & Lampitelli, N. (2017). *Stress in English long verbs: poster presented at the 25th Manchester phonology meeting, 25.–27.5.2017*, Manchester.

Daelemans, W., Gillis, S., & Durieux, G. (1994). The acquisition of stress: a data-oriented approach. *Computational Linguistics*, *20*(3), 421–451.

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

Deschamps, A. (1994). *De l'écrit à l'oral et de l'oral à l'écrit*. Paris: Ophrys.

Divjak, D., Milin, P., Ez-zizi, A., Józefowski, J., & Adam, C. (2021). What is learned from exposure: an error-driven approach to productivity in language. *Language, Cognition and Neuroscience*, *36*(1), 60–83. https://doi.org/10.1080/23273798.2020.1815813. Publisher: Routledge.

Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *Journal of Comparative Germanic Linguistics*, *17*, 59–96.

Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., & Goude, Y. (2021). Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, *116*(535), 1402–1412. https://doi.org/10.1080/01621459.2020.1725521. Publisher: Taylor & Francis.

Fudge, E. C. (1984). *English word-stress*. London: George Allen & Unwin.

Giegerich, H. (1999). *Lexical strata in English: morphological causes, phonological effects*. Cambridge: Cambridge University Press.

Giegerich, H. J. (1992). The limits of phonological derivation: spelling pronounciation and Schwa in English. *Linguistische Berichte*, *142*, 413–436.

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model.

Graves, A., Mohamed, A-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. 1303.5778 [cs].

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, *18*(5), 602–610. http://www.sciencedirect.com/science/article/pii/S0893608005001206.

Guierre, L. (1979). *Essai sur l'accentuation en anglais contemporain: Eléments pour une synthèse*, Université Paris-VII.

Hammond, M. (1999). *The phonology of English: a prosodic optimality-theoretic approach*. Oxford, New York: Oxford University Press.

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry*, *13*(2), 227–276.

Hoppe, D. B., Hendriks, P., Ramscar, M., & van Rij, J. (2022). An exploration of error-driven learning in simple two-layer networks from a discriminative learning perspective. *Behavior Research Methods*.

Hoppe, D. B., van Rij, J., Hendriks, P., & Ramscar, M. (2020). Order matters! Influences of linear order on linguistic category learning. *Cognitive Science*, *44*(11), e12910. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7685149/.

Jarmulowicz, L. (2006). School-aged children's phonological production of derived English words. *Journal of Speech, Language, and Hearing Research*, *49*, 294–308.

Jarmulowicz, L., Taran, V. L., & Hay, S. E. (2008). Lexical frequency and third-graders' stress accuracy in derived English word production. *Applied Psycholinguistics*, *29*(02), 213–235.

Jones, D. (2006). *Cambridge English pronouncing dictionary: edited by Peter Roach, James Hartman & Jane Setter*. Cambridge: Cambridge University Press.

Kager, R. (2012). Stress in windows: language typology and factorial typology. *Lingua*, *122*(13), 1454–1493.

Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior* (pp. 9–31). Miami: Miami University Press.

Kapatsinski, V. (2018). *Changing minds changing tools: from learning theory to language acquisition to language change*. Cambridge: MIT Press.

Kiparsky, P. (1982a). Lexical morphology and phonology. In I.-S. Yang (Ed.), *Linguistics in the morning calm: selected papers from SICOL* (pp. 3–91). Seoul: Hanshin.

Kiparsky, P. (1982b). Lexical phonology and morphology. *Linguistics in the Morning Calm*.

Kiparsky, P. (2015). Stratal OT: a synopsis and FAQs. In Y. E. Hsiao & L. H. Wee (Eds.), *Capturing phonological shades* (pp. 2–44). Newcastle upon Tyne: Cambridge Scholars Publishing.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148.

Linke, M., & Ramscar, M. (2020). How the probabilistic structure of grammatical context shapes speech. *Entropy*, *22*(1), 90. Publisher: Multidisciplinary Digital Publishing Institute.

McCully, C. (2003). Left-hand word-stress in the history of English. In P. Fikkert & H. Jacobs (Eds.), *Development in prosodic systems* (pp. 349–393). Berlin/Boston: de Gruyter.

Milin, P., Divjak, D., & Baayen, R. H. (2017a). A learning perspective on individual differences in skilled reading: exploring and exploiting orthographic and semantic discrimination cues. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(11), 1730–1751.

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017b). Discrimination in lexical decision. *PLoS ONE*.

Montgomery, S. (2001). *The case for synchronic orthographic primacy: the effect of literacy on phonological processing*. Edinburgh: University of Edinburgh.

Montgomery, S. (2005). Lax vowels, orthography and /ə/: the need for orthographic primacy. *Linguistische Berichte*, *201*, 14–64.

Moore-Cantwell, C. (2016). *The representation of probabilistic phonological patterns: Neurological, behavioral, and computational evidence from the English stress system*. Ph.D. dissertation, University of Massachusetts, Amherst.

Newell, H. (2021). Deriving Level 1/Level 2 affix classes in English: floating vowels, cyclic syntax. *Acta Linguistica Academica*.

Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. In M. I. Jordan, Y. LeCun, & S. A. Solla (Eds.), *Advances in neural information processing systems. Neural information processing systems conferences from 1988 to 1999 (CDROM)* (pp. 841–848).

Nieder, J., Tomaschek, F., Cohrs, E., & de Vijver, R. v. (2022a). Modelling Maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience*, *37*(3), 381–402. https://doi.org/10.1080/23273798.2021.1977835. Publisher: Routledge.

Nieder, J., van de Vijver, R., & Tomaschek, F. (2022b). "All mimsy were the borogoves" – a discriminative learning model of morphological knowledge in pseudo-word inflection. *Language, Cognition and Neuroscience*. Publisher: PsyArXiv.

Nixon, J. S. (2020). Of mice and men: speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, *197*, 104081. http://www.sciencedirect.com/science/article/pii/S0010027719302549.

Nixon, J. S., & Tomaschek, F. (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, pp. 585–591). https://cognitivesciencesociety.org/cogsci20/papers/0105/index.html.

Nixon, J. S., & Tomaschek, F. (2021). Prediction and error in early infant speech learning: a speech acquisition model. *Cognition*, *212*, 104697.

Pater, J. (2000). Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology*, *17*, 237–274.

Pearl, L., Ho, T., & Detrano, Z. (2016). An argument from acquisition: comparing English metrical stress representations by how learnable they are from child-directed speech. *Language Acquisition*, *24*(4), 307–342.

Ramscar, M. (2021). A discriminative account of the learning, representation and processing of inflection systems. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2021.2014062. Publisher: Routledge.

Ramscar, M., Dye, M., & Klein, J. (2013a). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017–1023.

Ramscar, M., Dye, M., & McCauley, S. (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.

Ramscar, M., Dye, M., Popick, H., & O'Donnell-McCarthy, F. (2011). The enigma of number: why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE*, *6*(7), e22501. Publisher: Public Library of Science.

Ramscar, M., & Yarlett, D. (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.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957.

Rescorla, R. (1988). Pavlovian conditioning – it's not what you think it is. *The American Psychologist*, *43*(3), 151–160.

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–69). New York: Appleton Century Crofts.

Saito, M., Tomaschek, F., & Baayen, H. (2020). Relative functional load determines co-articulatory movements of the tongue tip. In *12th international seminar on speech production (online conference)*, Providence, Rhode Island, USA (pp. 210–213).

Schmitz, D., Plag, I., Baer-Henney, D., & Stein, S. D. (2021). Durational differences of word-final /s/ emerge from the lexicon: modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology*, *12*, 2983. Publisher: Frontiers.

Shafaei-Bajestan, E., & Baayen, R. H. (2018). Wide learning for auditory comprehension. In *Interspeech* (pp. 966–970).

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423, 623–656.

Siegel, D. (1974). *Topics in English Morphology*. PhD. dissertation, MIT.

Stanton, J., & Steriade, D. (2014). *Stress windows and base faithfulness in English suffixal derivatives. Paper presented at the 22. Manchester phonology meeting (29.–31.5.2014).*

Terry, J., Ong, J. H., & Escudero, P. (2015). Passive distributional learning of non-native vowel contrasts does not work for all listeners. In *ICPhS*.

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (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.

Tomaschek, F., & Ramscar, M. (2022). Understanding the phonetic characteristics of speech under Uncertainty – implications of the representation of linguistic knowledge in learning and processing. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/article/10.3389/fpsyg.2022.754395.

Trevian, I. (2007). Stress-neutral endings in contemporary British English: an updated overview. *Language Sciences*, *29*(2–3), 426–450.

Tucker, B. V., Sims, M., & Baayen, R. H. (2019). *Opposing forces on acoustic duration*. Technical report. Publisher: PsyArXiv. psyarxiv.com/jc97w.

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014). Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PLoS ONE*, *9*(10), e109806.

Wanrooij, K. E., et al. (2015). *Distributional learning of vowel categories in infants and adults*.

Wells, J. (2008). *Longman pronunciation dictionary cased for pack*, 3rd revised edn, Pearson Longman.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, *21*(4), 221–226.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In *1960 WESCON convention record part IV*, New York (pp. 96–104). https://www.bibsonomy.org/bibtex/24c3b6ae932deb6bb1d04ad76c9c94a69/schaul.

Zamma, H. (2012). *Patterns and categories in English suffixation and stress placement: a theoretical and quantitative study*. PhD dissertation, University of Tsukuba.