



Theresa Matzinger* and Nikolaus Ritt

Phonotactically probable word shapes represent attractors in the cultural evolution of sound patterns

<https://doi.org/10.1515/cog-2021-0087>

Received July 24, 2021; accepted April 3, 2022; published online April 29, 2022

Abstract: Words are processed more easily when they have canonical phonotactic shapes, i.e., shapes that are frequent both in the lexicon and in usage. We explore whether this cognitively grounded constraint or preference implies testable predictions about the implementation of sound change. Specifically, we hypothesise that words with canonical shapes favour, or ‘select for’, sound changes that (re-) produce words with the same shapes. To test this, we investigate a Middle English sound change known as Open Syllable Lengthening (OSL). OSL lengthened vowels in disyllables such as ME /ma.kə/ *make*, but more or less only when they became monosyllabic and when their vowels were non-high. We predict that word shapes produced by this implementation pattern should correspond to the shapes that were most common among morphologically simple monosyllables and disyllables at the time when OSL occurred. We test this prediction against Early Middle English corpus data. Our results largely confirm our prediction: monosyllables produced by OSL indeed conformed to the shapes that were most frequent among already existing monosyllables. At the same time, the failure of OSL to affect disyllables (such as *body*) prevented them from assuming shapes that were far more typical of morphologically complex word forms than of simple ones. This suggests that the actuation and implementation of sound changes may be even more sensitive to lexical probabilities than hitherto suspected. Also, it demonstrates how diachronic data can be used to test hypotheses about constraints on word recognition and processing.

Keywords: cultural evolution; frequency effects; Middle English; Open Syllable Lengthening; sound change

***Corresponding author: Theresa Matzinger**, Department of English, University of Vienna, Vienna, Austria; and Centre of Language Evolution Studies, University of Toruń, Toruń, Poland, E-mail: theresa.matzinger@univie.ac.at. <https://orcid.org/0000-0001-5414-7962>

Nikolaus Ritt, Department of English, University of Vienna, Vienna, Austria, E-mail: nikolaus.ritt@univie.ac.at. <https://orcid.org/0000-0002-2253-5861>

1 Introduction

This paper deals with the hypothesis that the cultural evolution of sound patterns is constrained by a preference for probable word shapes because their relative frequency makes them predictable and easier to identify and process. This implies that word shapes that are in the majority, should be selected for and get even more frequent, whereas word shapes that are in the minority should be selected against and get less frequent.

1.1 The role of majority patterns in the cultural evolution of sound patterns

Our starting point is the observation that speakers are sensitive to the probability of phonotactic patterns in the lexical inventory (e.g., Blevins 2009; Divjak 2019; Mailhammer et al. 2015; Wedel 2006). This means that they classify words as instantiations of different relatively abstract shape types (for instance based on their syllable count, the weight of their syllables, or their stress patterns), and can sense how frequently different types of shape are instantiated among types of words in the lexicon and among word tokens in speech. Being able to distinguish between more and less probable word shapes, they can then exploit this difference in perception and production (e.g., Baayen et al. 2016; Gibson et al. 2019).

Word forms whose shapes are probable in the sense just described are identified more easily (Kelley and Tucker 2017), processed and repeated more quickly (Vitevitch and Luce 1998, 1999, 2005), learnt and memorized more easily (Storkel 2001; Storkel and Hoover 2010; Vitevitch et al. 2012), and produced more accurately (Goldrick and Larson 2008; Stemberger 2004; Vitevitch and Sommers 2003) than less frequent and less probable forms. Since word forms that are more easily recognized, repeated, memorized, and used, will also be transmitted more easily, word forms with more probable shapes ought to be historically more stable than word forms with rare or exceptional shapes. Thus, word form shapes that happen to be probable in a specific language will select for word form variants that conform to them, and function as attractors in phonological evolution (Blevins 2006, 2009; Blust 2007). Therefore, they ought to favour sound changes that (re-) produce them and thus stabilize or even bolster their own majority. This will further increase the advantages they afford in terms of word recognition, processing, and use. Crucially, this ought to be the case even if preferences for majority patterns are weak, since language history reflects the outcome of massively parallel and iterated transmission processes, which can amplify even very weak biases (e.g., Fehér et al. 2016; Kirby et al. 2014; Reali and Griffiths 2009; Smith et al. 2017).

As to biological and cognitive mechanisms that might favour more frequent phonological patterns over rarer ones, various proposals have been made. For example, motor entrenchment may reinforce the most frequent motor routines, articulatory movements, or neural pathways involved in speech production (Bybee 2002; Wedel 2007). At the same time, also speech perception may be biased towards shapes close to the centres of perceptual categories (cf. the ‘perceptual magnet effect’; Kuhl 1991). In combination, easier production and perception of frequent forms are apt to reinforce each other in a positive feedback loop (Pierrehumbert 2001, 2003; Wedel 2007, 2012; Wedel and Fatkullin 2017). Thus, there are mechanisms in both production and perception that are likely to make frequent forms easier to acquire, to use, and to transmit than less frequent ones.

The idea that frequent phonotactic patterns should act as attractors in language evolution is not new. In fact, it figures centrally in attempts to explain, in terms of cultural language evolution, why the shapes that words assume in natural languages tend to exploit only limited regions of the design space of logically possible or even physiologically viable forms. Also, lexical design spaces are typically more crowded in some phonotactic regions and less densely populated in others. That is to say, some phonotactic patterns occur more often than would be expected on the basis of mere chance, while others occur less often. This has been shown for several languages, for example in a synchronic study by Dautriche et al. (2017). Also, phonotactically similar words tend to be more semantically similar to one another than expected by chance (Dautriche et al. 2017; Monaghan et al. 2014; Tamariz 2004, 2008). In the face of such evidence, the view has gained increasing recognition that the attested distributions of word shapes in lexical design space result from evolutionary pressures that drive languages towards equilibria in which words are similar to one another without compromising distinctiveness.¹

¹ While efficiency in word perception, learning and use are well known factors in language evolution and change shapes (see e.g. Fedzechkina et al. 2012; Gibson et al. 2019), and may select for words with canonical shapes, there are also counteracting pressures that limit the extent to which word forms can become similar to one another (e.g., Kirby et al. 2015). Most prominent among them is the need for expressivity and contrastive patterns. As word form shapes become increasingly alike, they also become increasingly difficult to distinguish, which reduces their capacity to signal semantic contrasts as well (e.g., Dautriche et al. 2017; Tamariz 2004). For effects of this pressure on sound change see e.g., Blevins and Wedel (2009). Another factor that counteracts pressures on words to conform to canonical patterns is token frequency. High-frequency items are well known to resist regularisation and to be involved in the establishment of non-canonical patterns (such as the clusters /vr/, /mr/, or /ml/ in *every*, *memory*, and *family*; see already Bybee (2001).

Support for the hypothesized dynamics has been found in various areas. It comes, for example, from computer simulations. For instance, Wedel (2006, 2007, 2011) provides evidence of the attraction effect of frequent patterns by showing that positive production-perception feedback loops can increase their majority and create regularities in the lexicon. In addition, supporting evidence has been found in historical language studies such as Blevins (2004, 2006, 2009). A particularly pertinent set of developments in Austronesian is described in Blust (2007). It shows that disyllabic word forms, which were highly frequent (cf. also Kelso 1995), were likely to have been involved as attractors in the actuation of at least three different sound changes, namely initial vowel epenthesis, laryngeal loss, and loss of unstressed vowels between identical consonants. All of them ‘conspired’² to produce new disyllables and thereby bolstered their majority among lexical bases.

The present study therefore follows a line of reasoning that has been productively pursued in much recent research and applies it to a phenomenon that has so far not been approached from this perspective. Specifically, we focus on the lexical implementation of a Middle English sound change known as Open Syllable Lengthening (henceforth OSL; Luick 1964; Minkova 1982; Ritt 1994). OSL lengthened vowels in open syllables, produced forms such as Late Middle English /ma:k/ *make* or /be:vər/ *beaver* from earlier /ma.kə/ and /be.vər/, but had a large number of ‘exceptions’ (such as *body*, *copper*, *hammer*, and many others). We analyse a large set of corpus data with rigorous statistical methods, and show that the change was regularly implemented when the words resulting from it conformed to majority patterns, while it was implemented only sporadically in words where it would have produced minority patterns. Thus, apart from the fact that our results throw light on the puzzling implementation pattern of Middle English OSL, they provide further evidence in favour of the hypothesis that the actuation and implementation of sound changes may be conditioned by the statistical distribution of word-shapes among word types in the lexicon and among word tokens attested in speech.

1.2 The test case: Middle English Open Syllable Lengthening

OSL occurred between the thirteenth and fourteenth centuries and lengthened short vowels in stressed open syllables in disyllabic words, such as *make* (/ma.kə/ > /ma:k/), *hope* (/hɔ.pə/ > /hɔ:p/), *name* (/na.mə/ > /na:m/), or *beaver* (/be.vər/ > /be:və:r/; Bermúdez-Otero 1998b; Lahiri and Drescher 1999;

² In fact, the idea that sound changes can ‘conspire’ to implement canonical patterns can already be found in Jakobson (1929).

Mailhammer et al. 2015; Minkova 1982; Minkova and Lefkowitz 2020; Ritt 1994). However, there were crucial restrictions on its implementation. First, OSL did not regularly affect words with high vowels, such as *sin* (/si.nə/ > */si:n/).³ Second, the implementation of OSL depended on the phonotactic structure of its potential inputs: it only affected those disyllabic words consistently that became monosyllabic through the loss of their final syllable (a change known as schwa loss, as in /ma.kə/ > /ma:k/), which occurred roughly at the same period as OSL (Table 1; Minkova 1991, 2022; Minkova and Lefkowitz 2020). In contrast, it affected words only rarely if they remained disyllabic. Most of the few stable disyllables that were lengthened had sonorants in their second syllable (e.g., *beaver*, *bacon*), and a single one an obstruent (*naked*). Other stable disyllables – like *body* or *many* – were never lengthened (Table 1). Thus, the only items in which the change was implemented nearly categorically were disyllables like *make*, *name* or *hope*, which had non-high vowels and became monosyllabic due to final schwa loss.

A question that has intrigued historical linguists about this implementation pattern is why OSL was implemented primarily in words where the conditions that motivated the lengthening in the first place were lost. After all, when OSL was completed, the vowels in lengthened *make* /ma:k/, *name* /na:m/ or *hope* /hɔ:p/ were no longer in open syllables at all. The conclusion that has been drawn from this is that the lengthening must have been compensatory, i.e., that it made up for the weight loss induced by schwa loss (Bermúdez-Otero 1998a; Minkova 1982; Minkova and Lefkowitz 2020). This hypothesis receives support from the fact that stable disyllables in which schwa loss was at least optional (e.g., *beaver* could be realized as /be:vɪ/ or /be:vəɪ/) were also occasionally lengthened

Table 1: Implementation patterns of OSL, depending on the phonotactic structure of words (C = consonant, V = vowel, . = syllable boundary, ə = schwa, R = sonorant, T = obstruent; reproduced from Minkova and Lefkowitz 2020).

Phonotactic structure	Examples	Percentage of implementation
CV.Cə	<i>make, hope, name</i>	95.1%
CV.CVR	<i>beaver, hammer</i>	22.9%
CV.CVT	<i>habit, naked</i>	2.5%
CV.CV	<i>body, many</i>	0%

³ Some rare exceptions that were lengthened despite their high vowels are *door* (/du.rə/ > /do:r/), *beetle* (/bi.təl/ > /be:təl/), or *evil* (/y.vəl/ > /e:vəl/).

(Table 1). While this is descriptively adequate, it still raises the question of why compensation should have occurred, because the existence of forms such as *man* /man/ and *god* /gɔd/ suggests that realizations with short vowels such as */mak/, */nam/ and */hɔp/ would have been just as viable as the long realizations /ma:k/, /na:m/ and /hɔ:p/.

2 Hypotheses and predictions

2.1 General hypotheses

We hypothesise that lengthening was favoured in words of the *make*-type but not (or much less so) in the *beaver/habit/body*-type because this reflected frequencies of long versus short vowels in concurrent mono- and disyllabic words. The hypothesis that the implementation of OSL in Middle English may have been sensitive to the frequencies of phonotactic patterns at the time is not completely new either. A first version of the idea was proposed by Mailhammer (2010) and subjected to a quantitative investigation in Mailhammer et al. (2015). The study showed that, in comparison to other West Germanic Languages, Old English contained relatively more closed syllables with short vowels and relatively fewer open syllables with long vowels. It suggested that this might explain why OSL was also implemented less widely in English than in High German, Low German, or Frisian. Although we focus on pattern frequency too, our study differs from Mailhammer's in three ways. First, we count not only syllable types *per se*, but ask how frequent they were in exponents of simple or complex word forms. Second, we count syllable types separately for monosyllabic words and for disyllabic ones. Finally, we also distinguish syllables in terms of vowel height. This allows us to make much more specific predictions: for instance, since OSL was implemented categorically in words that became monosyllabic and had non-high vowels, we predict that at the time when the change set in, long non-high vowels should have been more frequent than short non-high vowels also among already existing monosyllables like *God* or *doom*. Since OSL was implemented only sporadically among morphologically simple disyllables that remained disyllabic, on the other hand, we expect that also among existing simple disyllables (like *āpum* 'son-in law' or *body*) long vowels should have been less frequent than short ones, while they may have been more frequent in complex disyllables (like *doom-es* 'doom.GEN' or *god-es* 'God.GEN').

Should our predictions be borne out, then the implementation of OSL would have further increased the frequency of shapes that were already typical – and therefore indicative – of morphologically simple words. It could therefore be explained as conditioned by (a) the frequency distribution of word form shapes in

the lexicon, and (b) a universal preference for that distribution to be skewed in favour of prototypical shapes.

2.2 Theoretical considerations and predictions for Early Middle English monosyllables

A fundamental question we had to address was which criteria we should apply to decide which existing Early Middle English word forms (Table 2b) should count as ‘like’ or ‘unlike’ OSL outputs (Table 2a). Of course, vowel length itself had to be the most decisive criterion, but vowel length distinguishes words not only on the segmental level, but also in terms of syllable weight. If one compares words in terms syllable weight, however, then OSL outputs such as lengthened /ma:k/ *make* are not only ‘like’ existing /mo:d/ *mood*, but also like /land/ *land*, because all three words have three segments in the rhyme. If one focusses only on vowel length, on the other hand, /ma:k/ is only ‘like’ /mo:d/, but not like /land/. Obviously, the level on which one compares words, will affect the number of existing forms that one finds to be shaped ‘like’ OSL outputs. In the following part, we explain how we dealt with that issue.

Regarding syllable weight, and treating final consonants as extrametrical,⁴ *god*-type items count as light, and *mood*-type and *land*-type items as heavy

Table 2: Phonotactic structure of (a) monosyllabic OSL outputs, and (b) already existing Early Middle English word forms that act as candidates for comparison (Recall that *CV was ungrammatical for major class items already in Middle English). Final consonants are counted as extrametrical.

	a. Lengthened OSL outputs	b. Existing Early Middle English monosyllabic word forms			
	<i>make</i> -type	<i>god</i> -type	<i>mood</i> -type	<i>land</i> -type	
Phonotactic shape	CVVC	CVC	CVVC	CVCC	CVCC
Pronunciation	/ma:k/	/gɔd/	/mo:d/	/land/	/gru:nd/
Syllable weight	heavy	light	heavy	heavy	
Vowel length	long	short	long	irrelevant for our classification	

⁴ We adopt this widely followed – albeit not uncontroversial – convention mainly because it allows us to refer to both ‘CVC’ in monosyllables and ‘CV’ in disyllables as ‘light’, and to all other types as ‘heavy’. We think this simplifies our terminology. None of our arguments depends on a specific theory of syllabification, however.

(Table 2b). Since OSL outputs (i.e., *make*-type items), count as heavy (Table 2a), they are ‘similar’ in that respect to all other heavy monosyllables. We would therefore predict that at the time when OSL and schwa loss set in, heavy monosyllables (*mood*-type and *land*-type items) should have been more frequent than light ones (*god*-type items; Figure 1, Prediction 1). If this was the case, the higher frequency of heavy syllables would have selected for lengthened, i.e., heavy, OSL outputs and against unlengthened, light competitor variants such as */mak/.

Applying a stricter similarity criterion, on the other hand, one would regard vowel length as the only relevant variable, and disregard items that are heavy just because they end in consonant clusters. Such a comparison considers only *god*-type and *mood*-type items and discards *land*-type items. On this stricter criterion, our hypothesis would be corroborated only if words with long vowels or diphthongs (i.e., words like *mood*) were more frequent in Early Middle English than words with short vowels (like *god*; Figure 1, Prediction 2). Clearly, this prediction is more difficult to meet than the first one, because the set of ‘light’ CVVC syllables is smaller than the set of all heavy syllables (i.e., CVVC, CVCC, CVVCC, etc.).

Finally, an even stricter similarity criterion would also take vowel height into account. Recall that high vowels were affected by OSL only very sporadically. Then our hypothesis would predict that among words with non-high vowels the *mood*-type should have been more frequent than the *god*-type, while among words with high vowels, the *lif*-type should not have been more frequent than the *cliff*-type (Figure 1, Prediction 3).

We decided to apply all three types of measure and to test three hypotheses. The weakest and most general one predicted the majority of monosyllables to be heavy; a stronger and more restricted one predicted that we should find more

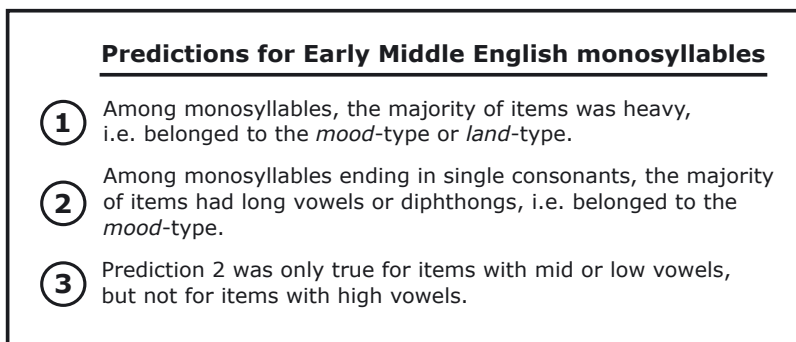


Figure 1: Summary of predictions for monosyllables.

CVVC items than CVC items; and the strongest of our hypotheses makes the most precise prediction, namely that we should find more CVVC items among words with non-high vowels, but not among words with high ones (Figure 1).

2.3 Theoretical considerations and predictions for Early Middle English disyllables

In the case of disyllables, we basically proceeded analogically (see Table 3). In one comparison, we classified them more generally in terms of the weight of their first syllables (holding light CV against heavy CVV, CVC, etc.). In the other comparison, we neglected all words with closed first syllables (i.e., CVC and heavier), and classified the remaining ones, i.e., syllables with open first syllables, in terms of their nuclei (holding CV syllables of the *mother*-type against CVV syllables of the *bailiff*-type, which had long vowels or diphthongs).⁵

Table 3: Phonotactic shape, syllable structure and syllable weight of the first syllable (σ_1) as well as examples of (a) disyllabic reflexes of OSL inputs and (b) already existing Early Middle English disyllables that act as candidates for comparison.

	a. Disyllabic reflexes of OSL inputs		b. Existing Early Middle English disyllabic word forms		
	lengthened (rare) <i>beaver</i> -type	unlengthened (frequent) <i>habit</i> -type	<i>mother</i> - type	<i>bailiff</i> - type	<i>finger</i> - type
Type of σ_1	CVV	CV	CV	CVV	CVC
Morphologically simple	bea.ver	ha.bit	mo.ther, bi.shop	bai.liff, ā.þum	fin.ger, mer.cy, an.gel
Morphologically complex	irrelevant for our study	irrelevant for our study	ta.l-es, no.m-en	doo.m-es	dren.ch-es, sor.h-en
Syllable structure	open	open	open	open	closed
Syllable weight	heavy	light	light	heavy	heavy
Vowel length	long	short	short	long	irrelevant for our classification

⁵ We did not consider vowel height, because the number of lengthened disyllables was too small for deriving any statistically significant observations.

Additionally, however, we also took the morphological structure of disyllables into account. Many of them, such as *doom-es* ‘doom.GEN’, *drench-es* ‘drink.PL’, or *sorh-en* ‘sorrow.PL’ were in fact morphologically complex. Outputs of OSL, however, were all morphologically simple. Thus, the question was not only how probable the shapes of OSL outputs were as exponents of words, but how probable they were as exponents of morphologically simple words. This is because morphologically simple word forms with shapes like complex ones invite unwarranted decomposition, and delay identification and processing (Post et al. 2008). This affects not only the ease with which word forms are acquired but also their historical stability (as has been shown in various studies in morphotactics such as Baumann and Kaźmierski 2018; Baumann et al. 2019; Calderone et al. 2014; Dressler and Dziubalska-Kołaczyk 2006; Korecky-Kröll et al. 2014; Ritt and Kaźmierski 2015). Therefore, morphologically simple disyllabic OSL outputs that were shaped like morphologically complex disyllables would not be identified and processed more easily because of that similarity. On the contrary, their similarity to complex disyllables would make them more difficult to identify and process, and would therefore select against them. This means that the frequency of complex items like *doom-es* ‘doom.GEN’ would not have motivated the selection of lengthened OSL outputs like *beaver*.

As OSL inputs were lengthened only sporadically if they remained disyllabic, we formulated the following predictions about disyllables that existed when OSL set in: among disyllables with light (and therefore also open) first syllables, the majority should have been morphologically simple (like *mo.ther*; see Figure 2, Predictions 1 & 3). Among syllables with any type of heavy first syllable (i.e., CVV, CVC, or heavier), the majority should have been morphologically complex (like *doo.m-es* or *dren.ch-es*; see Figure 2, Predictions 2 & 4). If this was the case, disyllables with light first syllables would have been typical and indicative of morphological simplicity, while disyllables with heavy first syllables would have signalled complexity. Thus, a reason why *habit*-type items were not affected by OSL could have been that this would have made it more difficult to identify them as morphologically simple.

There are, furthermore, two ways in which one can look at the correlation between phonotactic structure and morphological structure. On the one hand one can ask how likely it is that a specific type of phonotactic shape indicates either simplicity or complexity (Table 4a; and as outlined in the previous paragraph), and on the other hand one can ask how likely it is for simple or complex items to be represented by different types of phonotactic shape (Table 4b).

Both directions can affect the identification and the processing of words and their morphotactic structures. For the relationships in Table 4a, this is obvious: if

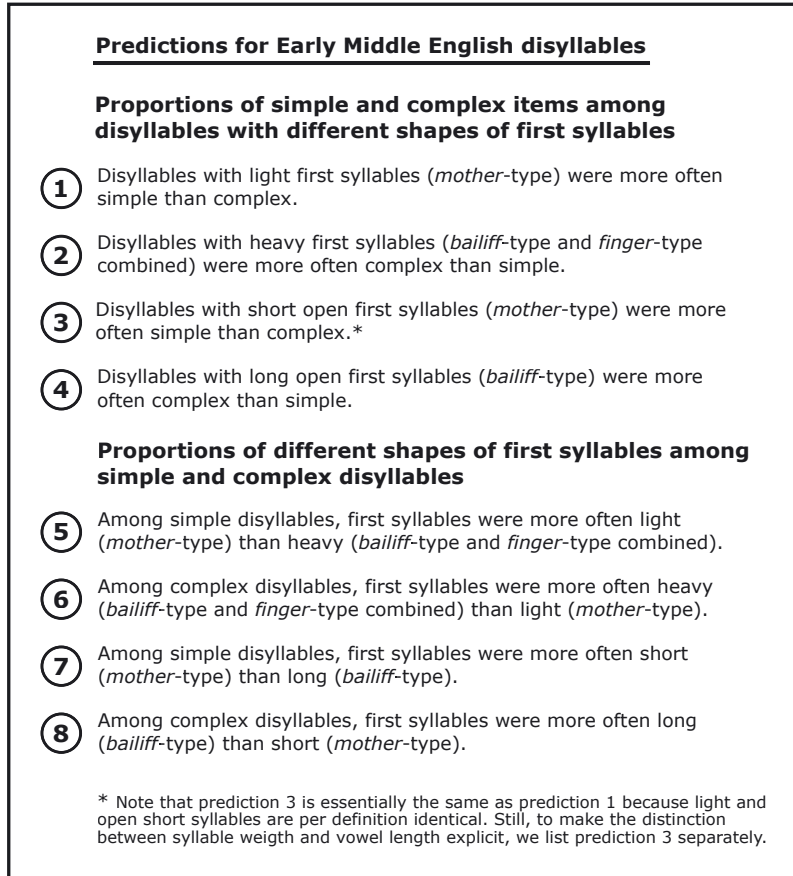


Figure 2: Summary of predictions for disyllables.

Table 4: Directionalities of correlations.

a.	What is encountered light first syllable (e.g., <i>mo.ther</i>) heavy first syllable (e.g., <i>doo.mes</i>)	→ What can be inferred → high probability of a simple word → high probability of a complex word
b.	What is expected simple item (e.g., { <i>mother</i> }) complex item (e.g., { <i>doom</i> } + { <i>es</i> })	→ What can be predicted → high probability of a light first syllable → high probability of a heavy first syllable

language users know that most items with light first syllables stand for simple words, this will help them to identify such shapes as simple words when they hear them. However, also the correlations in Table 4b are helpful for language processing. This is because perception is influenced by expectations (Cole et al. 2010; de Lange et al. 2018; McClelland and Elman 1986). For example, if context makes listeners expect a morphologically simple word (e.g., a noun such as *bishop*), and if they know that simple words are more likely to have light than heavy first syllables, it will be easier for them to perceive this word when it has indeed a light first syllable.

Therefore, our hypothesis would be supported most strongly if our data showed the expected correlations for all directionalities in Table 4. This means, we predicted not only that the majority of items with light first syllables should have been simple and the majority of items with a heavy first syllables complex (Table 4a; Figure 2, Predictions 1, 2, 3 & 4), but also that the majority of simple disyllables should have had light first syllables and the majority of complex disyllables should have had heavy ones (Table 4b; Figure 2, Predictions 5, 6, 7 & 8).

2.4 Type versus token frequencies

When counting frequencies of words with specific phonotactic patterns and morphological structures, we took both type and token frequencies into account. This is because the production, perception and processing of sound shapes may be influenced both by the number of different word form types, and the number of tokens, i.e., of utterances in which the types occur (Berg 2014). Type frequencies have been shown to be better predictors for phonological and morphological pattern learning than token frequencies (Baumann et al. 2019; Bybee 1995; Pierrehumbert 2016; Richtsmeier 2011 – but see Baumann and Kaźmierski 2018 for some counterevidence). We therefore expected to find more distinctive majority patterns among word form types than among tokens.

3 Methods

3.1 Data collection

To test if the frequencies of phonotactic shapes and morphological structures in Early Middle English word forms were as predicted, we used data from the LAEME corpus (The Linguistic Atlas of Early Middle English; Laing 2013). We chose

LAEME because it covers the period in which schwa loss and OSL began to unfold (1,150–1,325), and also because it is lemmatized and grammatically tagged at a high level of detail. The data were accessible in the form spelling type lists providing grammatical tags, the number of attestations/tokens of each type in the corpus, and the number of different texts in which they occur.

3.1.1 Monosyllables

For our data set of monosyllables, we extracted all word forms that were monosyllabic on the basis of their spelling (6,394 nouns, 8,809 verbs, 2,411 adjectives). This data needed additional processing: first, we excluded items with open syllables (such as *fe* ‘fee, livestock’, *dai* ‘day’, or *fa* ‘foe’) because these were not at all comparable to OSL outputs, which were all closed. Second, we excluded items whose coda was an inflectional suffix (such as *see* + *s* ‘sea.PL’, *sai* + *d* ‘say.PT’, *seo*+*ð* ‘see.3SGPRES’, or *ga* + *n* ‘go.INF’). Third, we excluded grammaticalized high frequency items because – due to their grammaticalization – they were not prototypical representatives of their word classes and were hypothesized not to serve as prototypical mental templates for newly emerging word shapes. The items excluded were the noun *man* (which also functioned as an indefinite pronoun), forms of *be*, *have*, or *do*, the modal verbs *may*, *will*, *shall*, *can*, the numeral adjectives *all*, *each*, *such*, *some*, and *which*. Finally, we excluded potential OSL outputs irrespectively of whether they were already spelt as monosyllables (such as *nom* ‘name’, *sac* ‘sake’, or *meet* ‘meat’) or whether they still had final <e> or other possibly silent vowel graphs. Our final dataset of monosyllables included 2,612 noun types, 1,606 verb types and 735 adjective types.

3.1.2 Disyllables

For our dataset of disyllables, we first extracted all disyllabic word forms (33,693 nouns, 39,642 verbs, 12,630 adjectives), and selected pseudo-random samples of 2,000 nouns, verbs and adjectives each. We made sure that the mix of items with high and low token frequencies in our samples reflected the token frequency mix in the complete dataset, except that we excluded *hapax legomena*. From our samples, we then excluded remaining items with transcription errors and items with unclear morphological structure, syllable weight or vowel length. Also, we excluded items whose final syllable was <-e>, since it was impossible to determine if in the target period (1,150–1,325) final *-e* was still pronounced as /e/, reduced to schwa, or already lost completely. Furthermore, its morphological status was unclear as well. The remaining dataset included 925 noun types, 1,134 verb types and 700 adjective types.

3.2 Data preparation and qualitative analysis

3.2.1 Preliminary remarks

The LAEME corpus provided us with lists of Early Middle English word forms attested in written texts. LAEME lists all spelling variants separately,⁶ and provides a lemma, a morpho-syntactic tag, and the token frequency for each variant. The phonological information we needed to derive from the written forms, was (a) syllable boundaries, (b) syllable weight, (c) the phonological length of vowels, and (d) the height of vowels if they were monophthongs. The morphological information that we needed to derive for disyllabic items was whether they were simple or complex.

Inevitably, our categorizations required a substantial degree of philological interpretation and were not always straightforward. This is because spelling never represents pronunciation faithfully, and Middle English spelling was particularly variable. Also, the large number of examples we had to characterize made it impossible to consider all aspects that a careful philological interpretation would normally require. Thus, not all our categorizations may stand up to close philological scrutiny. However, in cases where we found it difficult to decide between alternative interpretations, we tried to settle for the one that was less favourable to our predictions, to counteract the effects of a possible confirmation bias.

In the following, we describe and illustrate the basic principles we applied in our analysis. A more detailed discussion of the decisions we made during our categorizations, including further examples, can be found in the supplementary material.

3.2.2 Syllable boundaries and syllable weight

To determine syllable weight, we identified syllable boundaries (in the case of disyllables) and syllable codas. For that, we interpreted consonant graphs as representing phonological consonants more or less faithfully. We then assumed syllabification to be onset maximal. Thus, we would syllabify a form such as *knictes* ‘knight.PL’ as *knic.tes*, and a form like *bagges* ‘bag.PL’ as *ba.gges*.

On the basis of these syllabifications, we determined syllable weight. In the case of monosyllables, we counted all syllables as heavy that had more than a

⁶ Due to the absence of spelling standards, Middle English word forms could occur in a number of different spellings, and LAEME keeps them rigorously apart. Thus, a word form like *land* was attested in seven different monosyllabic spellings, namely <lond, long, lont, lond, land, loand, lonð>. Below (see Table 5) we explain how we dealt with this.

single coda consonant, irrespectively of the length of their vowel (e.g., *mauht* ‘might’ or *milc*, ‘milk’). In the case of disyllables, a single coda consonant in the first syllable counted as sufficient for making this syllable heavy (e.g., *knīc.tes* ‘knights’, *al.mes* ‘alms’, or *an.gel* ‘angel’).

3.2.3 Vowel quantity and vowel height

We determined vowel length and vowel height, i.e., vowel quality, by considering vowel length and quality in Old English or Modern English reflexes, and by consulting dictionaries such as the Oxford English Dictionary (<https://oed.com/>), the Middle English Dictionary (<https://quod.lib.umich.edu/m/middle-english-dictionary/dictionary>), or the Dictionary of Old English (<https://tapor.library.utoronto.ca/doe/>).

3.2.4 Morphological analysis

For the morphological analysis, we could rely on the grammatical tags provided in LAEME. For example, the form *comeð* ‘come’ is tagged as a second person plural imperative. Since the imperative has an evident phonological exponent, namely *-eð*, we confidently classified *comeð* as morphologically complex, and proceeded in the same way with all other cases.

3.3 Quantitative data analysis

To calculate the proportions of different phonotactic shapes and morphological patterns, we counted both type and token frequencies. Our basis for establishing what should count as a single type were unique combinations of sound shape, lemma and grammatical tag. For example, the seven spellings of *land* in Table 5

Table 5: Identification of types with regard to spelling, sound shape, lemma and grammatical tag.

Type	Spelling	Sound shape	Lemma	Tag
1	<i>lond</i>	CVCC	<i>land</i>	n
	<i>long</i>	CVCC	<i>land</i>	n
	<i>lont</i>	CVCC	<i>land</i>	n
2	<i>lond</i>	CVCC	<i>land</i>	n < pr
	<i>land</i>	CVCC	<i>land</i>	n < pr
	<i>loand</i>	CVCC	<i>land</i>	n < pr
	<i>lonð</i>	CVCC	<i>land</i>	n < pr

counted as two different types because – even though they shared the same sound shape and represented the same lemma – three spellings represented the nominative form, and four the oblique form ($n > pr$, i.e. noun forms preceded by prepositions). For token frequencies, we used the ones reported in LAEME.

To compare the proportions of phonotactic shapes and to establish which of them represented the majority, we calculated 95% confidence intervals. Confidence intervals that do not overlap with one another indicate significant differences between groups. Additionally, confidence intervals that do not include the 50% mark indicate that a pattern is either in the majority (above 50%), or in the minority (below 50%; Cumming 2012, 2014; Cumming and Finch 2005).

For disyllables, we additionally operationalized the relationship between morphological structure and sound shapes by calculating chi-squared tests and phi correlation coefficients, which measure the correlation between two binary variables (Everitt and Skrondal 2010; Warrens 2008; Yule 1912). A phi coefficient of 1 indicates a perfect correlation between morphological structure and sound shapes. This would be the case, for example, if all morphologically complex word forms had long vowels in their first syllables and all morphologically simple word forms short vowels, or *vice versa*. A phi coefficient of 0 indicates that there is no relationship between morphological structures and sound shapes and that listeners will be unable to infer morphological structure from sound shapes or *vice versa*. Commonly, phi coefficients around 0.3 indicate medium correlations and phi coefficients around 0.5 strong correlations (Cohen 1992). All calculations were done in *R* (version 3.6.0; R Development Core Team 2018).

4 Results: monosyllables

4.1 Syllable weight

Our analyses revealed that the proportions of heavy Early Middle English monosyllabic nouns, verbs and adjectives were clearly above 50%. This was true for word form types (nouns: 81.16%, verbs: 83.75%, adjectives: 81.23%; CIs do not include 50%; Figure 3a) and for word tokens (nouns: 86.47%, verbs: 77.17%, adjectives: 86.72%; CIs do not include 50%; Figure 3b). This means that the clear majority of Early Middle English monosyllables was heavy. Therefore, heavy monosyllables were much more probable as representatives of monosyllabic words than light monosyllables, which matches Prediction 1 (Figure 1), our weakest prediction about monosyllables.

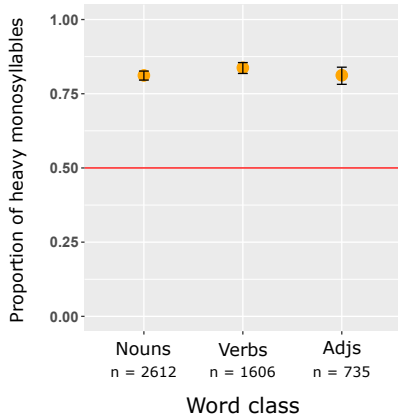
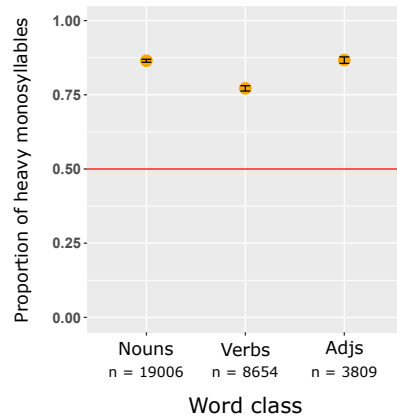
(a) Types**(b) Tokens**

Figure 3: Proportions of heavy monosyllable (a) types and (b) tokens in Early Middle English nouns, verbs and adjectives. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which sound patterns are maximally ambiguous.

4.2 Vowel length

Our analyses revealed that the clear majority of monosyllabic nouns, verbs and adjectives that ended in single consonants had long vowels or diphthongs, and only a minority had short vowels. Again, these relations held for word form types (nouns: 67.93%, verbs: 68.97%, adjectives: 70.07%; CIs do not include 50%; Figure 4a) and word tokens (nouns: 71.14%, verbs: 60.40%, adjectives: 78.29%; CIs do not include 50%; Figure 4b). Thus, words from the *mood*-type were more typical representatives of monosyllabic words than words from the *god*-type, which matches Prediction 2 (Figure 1), our stronger prediction about monosyllables.

4.3 Vowel length in high versus non-high monosyllables

A comparison between Early Middle English monosyllables with high and non-high vowels also revealed clear differences between these groups, which are roughly in line with Prediction 3 (Figure 1), our strongest prediction for monosyllables. The prediction is met unambiguously insofar as the clear majority of non-high vowels was long (types: nouns: 69.60%, verbs: 71.65%, adjectives: 74.73%; CIs do not include 50%; Figure 5a; tokens: nouns: 71.87%, verbs: 64.22%, adjectives: 79.06%; CIs do not include 50%; Figure 5b). However, our prediction

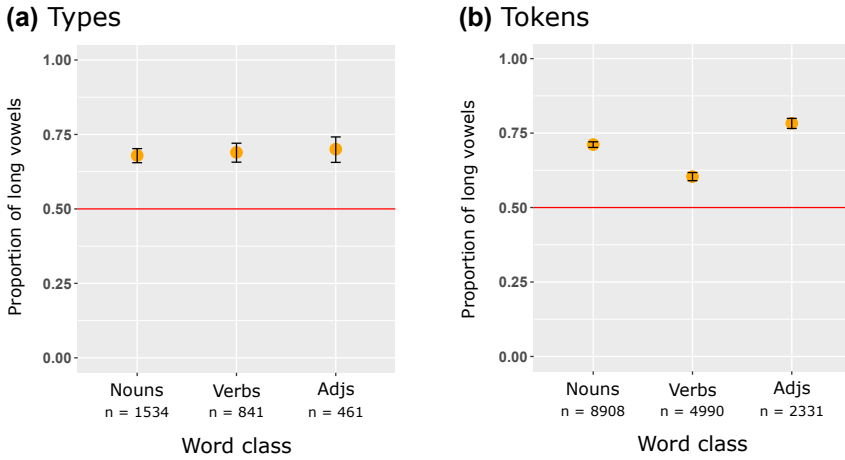


Figure 4: Proportions of monosyllable (a) types and (b) tokens with long vowels or diphthongs in Early Middle English nouns, verbs and adjectives. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which sound patterns are maximally ambiguous.

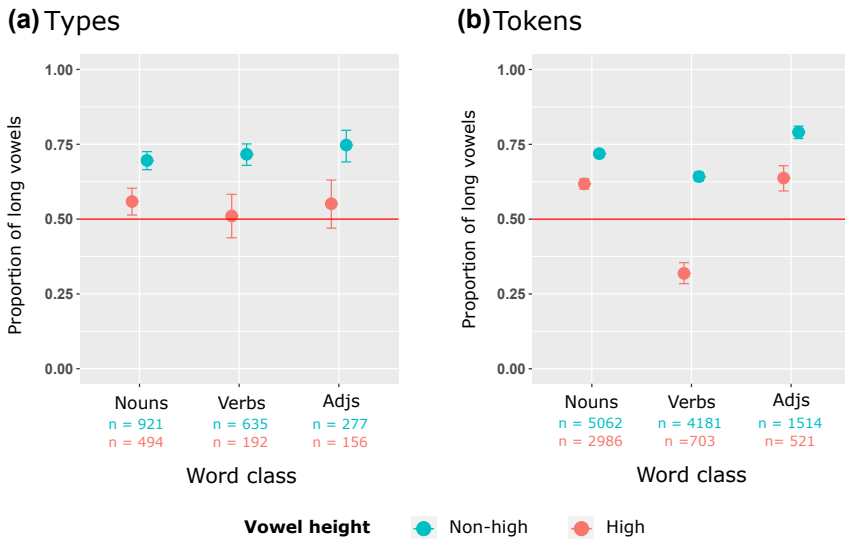


Figure 5: Proportions of monosyllable (a) types and (b) tokens with long vowels in Early Middle English nouns, verbs and adjectives, distinguished by vowel height. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which sound patterns are maximally ambiguous.

that the majority of high vowels would be short, was borne out only partly. On the type level, the proportion of short high vowels lay around 50% for verbs and adjectives, and in the case of nouns, a narrow majority of high vowels was in fact long (nouns: 55.87%, verbs: 51.04%, adjectives: 55.13%; CIs of verbs and adjectives include 50%; Figure 5a). On the level of tokens, the majority of vowels was short only for verbs, but not for nouns and adjectives (nouns: 61.79%, verbs: 31.86%, adjectives: 63.72%; Figure 5b). Nevertheless, the proportion of long vowels was always significantly greater among non-high vowels than among high ones (see non-overlapping confidence intervals in Figure 5 for noun, verb and adjective types and tokens), which is why, overall, there is some support for Prediction 3, but this support is only weak.

5 Results: disyllables

5.1 Proportions of simple and complex items among disyllables with different first syllable types

5.1.1 Syllable weight

Among disyllabic nouns and adjectives, the majority of word forms with light first syllables were morphologically simple. This was true both of types (nouns: 59.19%, adjectives: 64.35% simple; CIs do not include 50%; Figure 6a) and of tokens (nouns: 62.98%, adjectives: 85.18% simple; CIs do not include 50%; Figure 6b). In contrast, the majority of word forms with heavy initial syllables were complex.

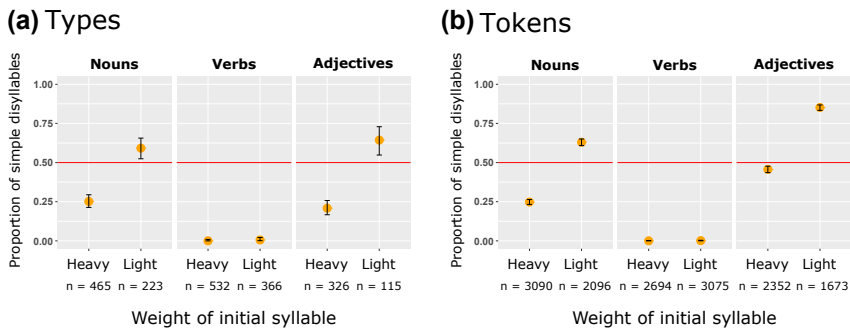


Figure 6: Proportions of morphologically simple word forms in Early Middle English disyllable noun, verb and adjective (a) types and (b) tokens with heavy and light initial syllables. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which sound patterns are least indicative for morphological structures.

This also held on the level of types (nouns: 25.16%, adjectives: 20.86% simple; CIs do not include 50%; Figure 6a) and tokens (nouns: 24.79%, adjectives: 45.62% simple; CIs do not include 50%; Figure 6b). The medium to strong correlations between morphological structure and initial syllable weight in nouns and adjectives (see results of chi-squared tests and phi-correlation coefficients in Tables 6 and 7) further support the significance of these relationships. – Thus, our Predictions 1 and 2 for disyllables (Figure 2) were borne out well among nouns and adjectives: disyllables with heavy open first syllables were more often complex than simple, and disyllables with light open first syllables were more often simple than complex.

Since our dataset did not include a sufficient number of morphologically simple verbs (two types and four tokens), no conclusions about verbs could be drawn.

5.1.2 Vowel length

As predicted, among disyllabic nouns and adjectives with short vowels in open first syllables, the majority were simple. This held among both types (nouns:

Table 6: Number of word form types with heavy and light initial syllables in morphologically simple and complex nouns, verbs and adjectives.

Word class	Weight	Simple	Complex	Correlation
Nouns	Heavy	117	348	$\chi^2 = 74.12, p < 0.001, \varphi = 0.33$
	Light	132	91	
Verbs	Heavy	0	532	$\chi^2 = 0.97, p = 0.324, \varphi = 0.03^a$
	Light	2	364	
Adjectives	Heavy	68	258	$\chi^2 = 71.67, p < 0.001, \varphi = 0.40$
	Light	74	41	

^aCorrelation of limited interpretability because of the low samples size of simple items in our dataset.

Table 7: Number of word tokens with heavy and light initial syllables in morphologically simple and complex nouns, verbs and adjectives.

Word class	Weight	Simple	Complex	Correlation
Nouns	Heavy	766	2,324	$\chi^2 = 755.85, p < 0.001, \varphi = 0.38$
	Light	1,320	776	
Verbs	Heavy	0	2,694	$\chi^2 = 1.88, p = 0.170, \varphi = 0.02^a$
	Light	4	3,071	
Adjectives	Heavy	1,073	1,279	$\chi^2 = 647.98, p < 0.001, \varphi = 0.40$
	Light	1,425	248	

^aCorrelation of limited interpretability because of the low samples size of simple items in our dataset.

59.19%, adjectives: 64.34% simple; CIs do not include 50%; Figure 7a) and tokens (nouns: 62.98%, adjectives: 85.18% simple; CIs do not include 50%; Figure 7b). These numbers are identical to those for disyllables with light first syllables (Figure 6) because light syllables are per definition identical to open short ones.

Among disyllables with long vowels in open first syllables, there was a difference between the type level and the token level. On the type level, both nouns and adjectives were distributed as we predicted: the majority of both adjectives and nouns with long vowels in their open first syllable were indeed complex (nouns: 18.65%, adjectives: 30.06% simple; CIs do not include 50%; Figure 7a). On the token level, however, adjectives differed from nouns. Among nouns, the majority of items with long vowels in open first syllables were complex (14.95% simple; CIs do not include 50%; Figure 7b). Among adjectives, however, the majority were simple (63.32% simple; CIs do not include 50%; Figure 7b), although that majority was not as great as among adjective tokens with short vowels in open first syllables. – In spite of the odd behaviour of adjective tokens, however, the relationships between short and long vowels in open first syllables and complex versus simple word forms displayed medium to strong correlations between morphological structure and initial vowel length in nouns and adjectives (see results of chi-squared tests and phi-correlation coefficients in Tables 8 and 9). Thus, our Predictions 3 and 4 for disyllables (Figure 2) were on the whole borne out well: disyllables with long vowels in open first syllables were more often complex than simple, and disyllables with short vowels in open first syllables were more often simple than complex. – Once again, it has to be pointed

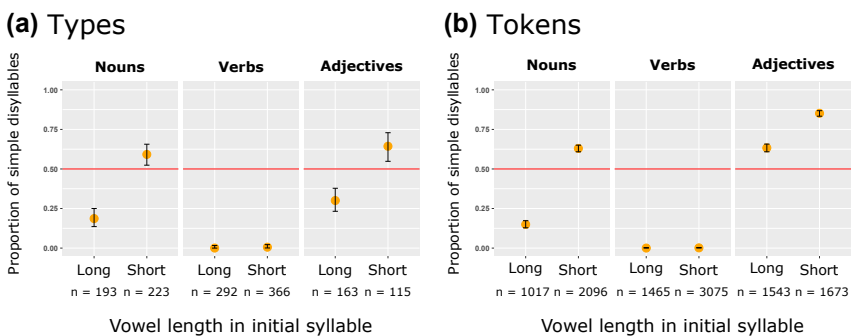


Figure 7: Proportions of morphologically simple word forms in Early Middle English disyllable noun, verb and adjective (a) types and (b) tokens with long and short initial syllables. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which sound shape patterns are least indicative for morphological structures.

Table 8: Number of word form types with long and short vowels in their initial syllables in morphologically simple and complex nouns, verbs and adjectives. Note that the data for items with short initial vowels is identical to the data for light initial syllables in Table 6.

Word class	Vowel length	Simple	Complex	Correlation
Nouns	Long	36	157	$\chi^2 = 68.95, p < 0.001, \varphi = 0.41$
	Short	132	91	
Verbs	Long	0	292	$\chi^2 = 0.30, p = 0.581, \varphi = 0.02^a$
	Short	2	364	
Adjectives	Long	49	114	$\chi^2 = 30.76, p < 0.001, \varphi = 0.33$
	Short	74	41	

^aCorrelation of limited interpretability because of the low samples size of simple items in our dataset.

Table 9: Number of word tokens with long and short vowels in their initial syllables in morphologically simple and complex nouns, verbs and adjectives. Note that the data for items with short initial vowels is identical to the data for light initial syllables in Table 7.

Word class	Vowel length	Simple	Complex	Correlation
Nouns	Long	152	865	$\chi^2 = 631.83, p < 0.001, \varphi = 0.45$
	Short	1,320	776	
Verbs	Long	0	1,465	$\chi^2 = 0.72, p = 0.398, \varphi = 0.01^a$
	Short	4	3,071	
Adjectives	Long	977	566	$\chi^2 = 201.71, p < 0.001, \varphi = 0.25$
	Short	1,425	248	

^aCorrelation of limited interpretability because of the low samples size of simple items in our dataset.

out that the low number of morphologically simple verb forms, did not allow us to draw any conclusions.

5.2 Proportions of different first-syllable shapes among simple and complex disyllables

5.2.1 Syllable weight

Among morphologically complex disyllabic adjectives and nouns, the clear majority of word forms had heavy first syllables. This was true both for types (nouns: 79.27%, adjectives: 68.29% heavy; CIs do not include 50%; Figure 8a) and for tokens (nouns: 74.97%, adjectives: 83.76% heavy; CIs do not include 50%; Figure 8b). Among morphologically simple disyllabic adjectives and nouns, the proportions of heavy first syllables were lower. For types, the proportions of heavy

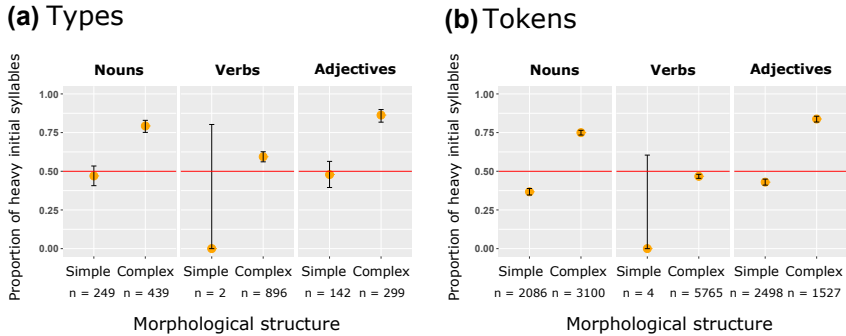


Figure 8: Proportions of disyllabic word forms with heavy initial syllables in morphologically simple and complex Early Middle English noun, verb and adjective (a) types and (b) tokens. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which morphological structures are least indicative for sound shapes.

first syllables were around 50% (nouns: 46.99%, adjectives: 47.89% heavy; CIs include 50%; Figure 8a) and for tokens, they were below 50% (nouns: 36.72%, adjectives: 42.95% heavy; CIs do not include 50%; Figure 8b). For nouns and adjectives, this relationship is also reflected in significant medium to strong correlations between the morphological structure of words and the weight of their first syllables (see results of chi-squared tests and phi-correlation coefficients in Tables 6 and 7).

Among disyllabic verbs, the majority of complex word form types (59.38% heavy; Figure 8a) had heavy first syllables. However, among complex verb tokens, the proportion of heavy first syllables was slightly below 50% (46.73% heavy; Figure 8b). About simple disyllabic verbs, nothing can be said because there were hardly any of them in our sample (two types, and four tokens, which is not surprising since verbal inflection was still intact in Early Middle English). Overall, our data match our Prediction 6 for disyllables (Figure 2) well: first syllables were more often heavy in complex word forms. Prediction 5 (Figure 2) is also met, but not as clearly: first syllables are indeed more often light in simple word form tokens, but for simple word form types, the proportion of light syllables lies just around 50%.

5.2.2 Vowel length

For disyllables with open first syllables, our results were very similar to those for syllable weight. Among complex disyllabic nouns and adjectives with open first syllables, the clear majority had long vowels in their first syllable. This was true

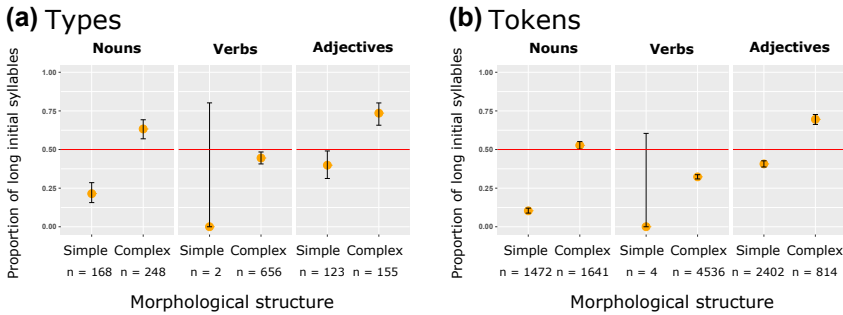


Figure 9: Proportions of disyllabic word forms with long initial syllables in morphologically simple and complex Early Middle English noun, verb and adjective (a) types and (b) tokens. Error bars represent 95% confidence intervals. The red line indicates 50%, i.e., proportions at which morphological structures are least indicative for sound shapes.

both for types (nouns: 63.31%, adjectives: 73.55% long; CIs do not include 50%; Figure 9a), and for tokens (nouns: 52.71%, adjectives: 69.53% long; CIs do not include 50%; Figure 9b). In contrast, the majority of simple word form types had short vowels in first syllables. Once again, this held for types (nouns: 21.43%, adjectives: 39.84% long; CIs do not include 50%; Figure 9a) and tokens (nouns: 10.32%, adjectives: 40.67% long; CIs do not include 50%; Figure 9b).

These relations also manifest in significant medium to strong correlations between morphological structure and vowel length in word-initial syllables (see results of chi-squared tests and phi-correlation coefficients in Tables 8 and 9). Thus, among nouns and adjectives, our Predictions 7 and 8 (Figure 2) were borne out well: Among complex disyllables, the majority of first syllables were long rather than short, while among simple disyllables, the opposite was true.

Once again, the picture is less clear for verbs. In contrast to nouns and adjectives, the majority of complex disyllabic verbs had short vowels in their first syllables, although for word form types, the proportion of short vowels in initial syllables lies only marginally above 50% (verb types: 44.51%; verb tokens: 32.30% long; Figure 9). Again, we cannot say anything about simple verbs.

6 Discussion

Our results provide support for the majority of our predictions: at the time when OSL set in, long and short vowels and heavy and light syllables were distributed among monosyllabic and disyllabic word forms so that the way in which OSL was implemented stabilized or even increased the probability of word form shapes that

were already in the majority. Thus, the regular lengthening of non-high vowels in the *make*-type increased the probability of monosyllabic word forms to be heavy rather than light, and to have long rather than short vowels if they did not end in consonant clusters. Likewise, the failure of high vowels to lengthen in such cases matches the fact that high vowels in Early Middle English monosyllables were typically not more often long than short.

Among disyllables, the failure of most vowels to undergo OSL corresponds to the fact that the majority of simple disyllables had short vowels in light first syllables at the time when OSL set in. In contrast, heavy first syllables (no matter of closed or open) were more frequent among complex disyllabic word forms. That relationship also held the other way round: if a disyllabic word had a light first syllable (i.e., a short vowel), then it would have been simple in the majority of cases, and if its first syllable was heavy it would have been complex. Thus, the distribution of long and short vowels among disyllables was a good indicator of their morphological structure. The implementation of OSL not only helped to maintain this relationship by not lengthening vowels in words of the *habit*-type, but it even increased that indicativeness further, albeit indirectly, by lengthening vowels in words that became monosyllabic. This is because any inflected forms of such words (e.g., *makeð* ‘make 3.SG.PRES’, or ‘*names* ‘name.PL’) would increase the already high probability of disyllables with heavy first syllables to be complex.

Of course, our study has been purely correlational. Since the correlations we have found are quite strong and quite specific, however, it seems worthwhile to discuss the causalities they might reflect. Like most sound changes, OSL is likely to have started on the phonetic level, by lengthening the duration of vowels that were phonologically short. Thus, their duration would have become ambiguous as an indicator of the intended short vowels, and these vowels may have been reinterpreted as reflecting phonologically long ones. So, phonologically long and short variants will have competed as phonological representations of OSL inputs.⁷ If words, and word forms with probable phonotactic shapes are easier to identify and to learn than words with less probable shapes, this may have selected for those variants that did have the more probable and morpho-syntactically more indicative shapes. Among words of the *make*-type, these were the variants with long vowels, and among words of the *beaver/habit*-types, they were the ones with short vowels. As far as we see it, such an account would be logically consistent, and there is much independent evidence for all the processes and preferences it needs to appeal to – both from socio-historical

⁷ Evidence of such variability comes also from the fact that OSL is reflected slightly differently in different English dialects. For example, Yorkshire English has a short vowel in *water* (Mailhammer et al. 2015).

phonology and from psycholinguistics. In particular, such an account would clearly be compatible with, and support the general hypothesis that a preference for word forms to assume probable shapes represents a possibly universal cognitive bias that may interact with other factors to constrain the evolution of sound patterns (Ambridge et al. 2015; Bybee 2007; Diessel 2007; Divjak 2019; Ellis 2002).

It also needs to be taken into consideration that the way in which preferences based on lexical statistics interact with other factors may be complex. Consider for example the case of high vowels in monosyllables. While their failure to implement OSL seems to be predictable by the fact that long items like *wif* 'wife' /wi:f/ or *house* /hu:s/ were not significantly more probable at the time when OSL set in than short items like *wit* or *full*, it may equally well have been caused by the inherently shorter duration of high vowels in comparison to non-high ones (Delattre 1962; House 1961; Lehiste 1970; Lisker 1974). Indeed, the inherent shortness may underlie both the relative rarity of *wif*-type words and the failure of high vowels to undergo OSL at the same time. However, even if that should be the case, the two factors may have mutually supported one another.

More generally speaking, the potential importance of lexical probabilities, which our findings suggest, does not invalidate the importance of other phonological factors. These include the open syllable condition itself, the quality of the postvocalic consonants, or the structure of the second syllable if it was retained. Since our focus has been on the potential impact of lexical probability, we have not discussed the details of these phonological conditions on OSL (see e.g., Bermúdez-Otero 1998b; Lahiri and Drescher 1999; Mailhammer et al. 2015; Minkova 1982, 2022; Minkova and Lefkowitz 2020; Ritt 1994 for in depth discussions). Therefore, our findings are clearly not intended to compete with extant accounts but rather to complement them.

A final aspect is that, overall, type level results were more strongly compatible with our predictions than token level results. This is plausible because it is compatible with similar insights on language acquisition and learning (e.g., Bybee 1995; Ellis 2002; Endress and Hauser 2011; Lieven 2010). The correlations we have demonstrated involve abstractions on a comparably high level, namely between syllabic structures that can be realized by a variety of different segment sequences, and morphotactic structures that can likewise be realized by a variety of different morpheme combinations. To learn that there is a statistical correlation between abstract phonotactic patterns such as an initial heavy syllable and abstract complex morphotactic patterns such as stem + suffix is very likely to require exposure to many different types of these patterns. A few types may not be enough, even if they are highly frequent in terms of tokens. Thus, the fact that type-level results show clearer correlations than token-level results is not surprising.

7 Limitations, conclusion, and outlook

Although practically all our predictions have been borne out, it needs to be stressed that they merely support the plausibility of the general hypothesis that lexical probabilities may constrain the implementation of sound changes. They do not prove it. Among other things, this is because our argumentation has been abductive. We have started from the observation that OSL was regularly implemented among disyllables that had non-high vowels and became monosyllabic, and that it was implemented only rarely among words with high vowels and among words that retained their second syllable. We then defined the conditions under which a preference for words with probable and morpho-syntactically indicative sound shapes would predict the attested implementation pattern, and finally we enquired if these conditions held. That we did indeed find the necessary conditions to hold, therefore merely suggests that our hypothesis is plausible but does not prove that the causalities it implies were really involved in producing the attested implementation pattern.

There are other limitations to our study. For instance, it could not do justice to dialectal diversity, even though OSL is reflected differently in different Modern English dialects. The example of *water*, which has a short vowel in Yorkshire English (Mailhammer et al. 2015) is just one case in point. On the other hand, we have treated our Early Middle English data without doing justice to the fact that they come from a heterogeneous set of different individual text languages, representing different varieties or even idiolects. Although this is common practice in studies of Early English, it needs to be acknowledged that caution is clearly warranted when one interprets quantitative findings derived from such data. Finally, we have not asked what effect the selection for canonical word form shapes may have had on their distinctiveness. Pressures in favour of lexical conformity are known to be counteracted by a need for lexical contrasts (Blevins and Wedel 2009; Dautriche et al. 2017; Tamariz 2004), and this might be relevant particularly in the case of monosyllabic words: not only has their number continued to rise during and after the Middle English period, but at the same time, various developments – such as vowel shortening in words like *dead* (< /dɛ:d/) – have increased the number of short monosyllables of the *god*-type, which were still rare at the time of OSL, as well. Although this might suggest that the lexical space occupied by – canonical – heavy monosyllables was getting over-crowded, we have not been able to address this question here and need to leave it for future research.

However, despite such limitations we take our findings to be interesting enough to warrant further research. In particular, and even though our results

concern only a single and quite specific case of a sound change, they suggest that lexical probabilities may play a greater role in the actuation and the implementation of phonological change than currently known. Given the increasing availability of digitized corpora and dictionaries of historical language stages, investigations of such a role may become more practicable than they have been and could also be extended to languages beyond English and phenomena beyond OSL. Such research could further support that sound changes are more likely to be actuated and implemented if they stabilize or increase the probability of already probable sound patterns, which would considerably advance our understanding of phonological evolution.

Data availability statement

The datasets generated and analysed during the current study (doi: <https://doi.org/10.17605/OSF.IO/CKMSH>) are available in the Open Science Framework repository and can be accessed at <https://osf.io/ckmsh/>.

Acknowledgements: We would like to thank Klaus Hofmann, Andrew Wedel, and two anonymous reviewers for valuable feedback on earlier drafts of this paper.

References

- Ambridge, Ben, Evan Kidd, Caroline F. Rowland & Anna L. Theakston. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42. 239–273.
- Baayen, R. Harald, Petar Milin & Michael Ramscar. 2016. Frequency in lexical processing. *Aphasiology* 30(11). 1174–1220.
- Baumann, Andreas & Kamil Kaźmierski. 2018. Assessing the effect of ambiguity in compositionality signaling on the processing of diphones. *Language Sciences* 67(May). 14–32.
- Baumann, Andreas, Christina Prömer & Nikolaus Ritt. 2019. Word form shapes are selected to be morphotactically indicative. *Folia Linguistica* 40(1). 129–151.
- Berg, Thomas. 2014. On the relationship between type and token frequency. *Journal of Quantitative Linguistics* 21(3). 199–222.
- Bermúdez-Otero, Ricardo. 1998a. Lexicon optimization: Irregular vowel length changes in Middle English. In *X International Conference on English Historical Linguistics Manchester*, vol. 22. Available at: <http://bermudez-otero.com/10icehl.pdf>.
- Bermúdez-Otero, Ricardo. 1998b. Prosodic optimization: The Middle English length adjustment. *English Language and Linguistics* 2(02). 169–197.
- Blevins, Juliette. 2004. *Evolutionary phonology – The emergence of sound patterns*, vol. 112. Cambridge: Cambridge University Press.

- Blevins, Juliette. 2006. A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics* 32(2). 117–166.
- Blevins, Juliette. 2009. Structure-preserving sound change: A look at unstressed vowel syncope in Austronesian. In Adelaar Alexander & Andrew Pawley (eds.), *Austronesian historical linguistics and culture history: A Festschrift for Bob Blust*, 33–49. Canberra: Pacific Linguistics.
- Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26(2). 143–183.
- Blust, Robert. 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. *Phonology* 24(1). 1–36.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14(3). 261–290.
- Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Calderone, Basilio, Chiara Celata, Katharina Korecky-Kröll & Wolfgang Ulrich Dressler. 2014. A computational approach to morphonotactics: Evidence from German. *Language Sciences* 46. 59–70.
- Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112(1). 155–159.
- Cole, Jennifer, Yoonsook Mo & Mark Hasegawa-Johnson. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2). 425–452.
- Cumming, Geoff. 2012. *Understanding the new statistics*. New York: Routledge.
- Cumming, Geoff. 2014. The new statistics: Why and how. *Psychological Science* 25(1). 7–29.
- Cumming, Geoff & Sue Finch. 2005. Inference by eye confidence intervals and how to read pictures of data. *American Psychologist* 60(2). 170–180.
- Dautriche, Isabelle, Kyle Mahowald, Edward Gibson, Anne Christophe & Steven T. Piantadosi. 2017. Words cluster phonetically beyond phonotactic regularities. *Cognition* 163. 128–145.
- Dautriche, Isabelle, Kyle Mahowald, Edward Gibson & Steven T. Piantadosi. 2017. Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science* 41(8). 2149–2169.
- Delattre, Pierre. 1962. Some factors of vowel duration and their cross-linguistic validity. *The Journal of the Acoustical Society of America* 34. 1141–1143.
- Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25. 108–127.
- Divjak, Dagmar. 2019. *Frequency in language – Memory, attention and learning*. Cambridge: Cambridge University Press.
- Dressler, Wolfgang U. & Katarzyna Dziubalska-Kolaczyk. 2006. Proposing morphonotactics. *Wiener Linguistische Gazette* 73(1–19). 108–109.
- Ellis, Nick C. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24. 143–188.
- Endress, Ansgar D. & Marc D. Hauser. 2011. The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *Journal of Experimental Psychology: Learning Memory and Cognition* 37(1). 77–95.

- Everitt, Brian & Anders Skrondal. 2010. *The Cambridge dictionary of statistics*, 4th edn. Cambridge: Cambridge University Press.
- Fedzechkina, Maryia, T. Florian Jaeger & Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America* 109(44). 17897–17902.
- Fehér, Olga, Elizabeth Wonnacott & Kenny Smith. 2016. Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language* 91. 158–180.
- Gibson, Edward, Richard Futrell, Steven T. Piandadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407.
- Goldrick, Matthew & Meredith Larson. 2008. Phonotactic probability influences speech production. *Cognition* 107(3). 1155–1164.
- House, Arthur S. 1961. On vowel duration in English. *The Journal of the Acoustical Society of America* 33(9). 1174–1178.
- Jakobson, Roman. 1929. *Remarks on the phonological evolution of Russian in comparison with the other Slavic languages*. Cambridge, MA: MIT press.
- Kelley, Matthew C. & Benjamin V. Tucker. 2017. The effects of phonotactic probability on auditory recognition of pseudo-words. *The Journal of the Acoustical Society of America* 141(5). 4038.
- Kelso, Scott J. A. 1995. *Dynamic patterns – The selforganization of brain and behavior*. Cambridge, MA: MIT Press.
- Kirby, Simon, Tom Griffiths & Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28. 108–114.
- Kirby, Simon, Monica Tamariz, Hannah Cornish & Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141(August). 87–102.
- Korecky-Kröll, Katharina, Wolfgang Ulrich Dressler, Eva Maria Freiberger, Eva Reinisch, Karlheinz Mörth & Libben Gary. 2014. Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences* 46. 48–58.
- Kuhl, Patricia K. 1991. Human adults and human infants show a ‘Perceptual Magnet Effect’ for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50(2). 93–107.
- Lahiri, Aditi & B. Elan Dresher. 1999. Open Syllable Lengthening in West Germanic. *Language* 75(4). 678–719.
- Laing, Margaret. 2013. *A linguistic atlas of Early Middle English, 1150–1325*. Edinburgh: The University of Edinburgh.
- Lange, Floris P. de, Micha Heilbron & Peter Kok. 2018. How do expectations shape perception? *Trends in Cognitive Sciences* 22(9). 764–779.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: MIT Press.
- Lieven, Elena. 2010. Input and first language acquisition: Evaluating the role of frequency. *Lingua* 120(11). 2546–2556.
- Lisker, Leigh. 1974. On ‘explaining’ vowel duration variation. *Glossa* 8(2). 233–246.
- Luick, Karl. 1964. *Historische Grammatik der Englischen Sprache* [Historical grammar of English], 1st edn. Leipzig: Tauchnitz.
- Mailhammer, Robert. 2010. Thoughts on the genesis and the development of syllable cut in English. *Anglia* 127(2). 261–282.

- Mailhammer, Robert, William W. Kruger & Makiyama Alexander. 2015. Type frequency influences phonological generalizations: Eliminating stressed open syllables with short vowels in West Germanic. *Journal of German Linguistics* 27(3). 205–237.
- McClelland, James L. & Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18(1). 1–86.
- Minkova, Donka. 1982. The environment for Open Syllable Lengthening in Middle English. *Folia Linguistica Historica* 3(1). 29–58.
- Minkova, Donka. 1991. *The history of final vowels in English: The sound of muting*, vol. 4. Berlin and New York: Mouton de Gruyter.
- Minkova, Donka. 2022. Preference theory and the uneven progress of degemination in Middle English. In Patrizia Noel Aziz Hanna & Laura C Smith (eds.), *Linguistic preferences (Trends in linguistics Studies and monographs 358)*, 265–292. Berlin: Mouton de Gruyter.
- Minkova, Donka & Michael Lefkowitz. 2020. Middle English Open Syllable Lengthening (MEOSL) or Middle English Compensatory Lengthening (MECL)? *English Language and Linguistics* 25(1). 1–26.
- Monaghan, Padraic, Richard C. Shillcock, Morten H. Christiansen & Simon Kirby. 2014. How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). 1–12.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul J. Hopper (eds.), *Frequency effects and the emergence of lexical structure*, 137–157. Amsterdam: Benjamins.
- Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46(2–3). 115–154.
- Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33–52.
- Post, Brechtje, William D. Marslen-Wilson, Billi Randall & Lorraine K. Tyler. 2008. The processing of English regular inflections: Phonological cues to morphological structure. *Cognition* 109(1). 1–17.
- R Development Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Realí, Florencia & Thomas L. Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111(3). 317–328.
- Richtsmeier, Peter T. 2011. Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology* 2(1). 157–183.
- Ritt, Nikolaus. 1994. *Quantity adjustment: Vowel lengthening and shortening in Early Middle English*. Cambridge: Cambridge University Press.
- Ritt, Nikolaus & Kamil Kaźmierski. 2015. How rarities like gold came to exist: On co-evolutionary interactions between morphology and lexical phonotactics. *English Language and Linguistics* 19. 1–29.
- Smith, Kenny, Perfors Amy, Olga Fehér, Samara Anna, Kate Swoboda & Elizabeth Wonnacott. 2017. Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 372(1711). 1–20.
- Stemberger, Joseph Paul. 2004. Neighbourhood effects on error rates in speech production. *Brain and Language* 90(1–3). 413–422.
- Storkel, Holly L. 2001. Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research* 44(6). 1321–1337.

- Storkel, Holly L. & Jill R. Hoover. 2010. An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods* 42(2). 497–506.
- Tamariz, Monica. 2004. Exploring the adaptive structure of the mental lexicon. University of Edinburgh PhD thesis. <http://www.isrl.uiuc.edu/~amag/langev/paper/tamariz05phd.html>.
- Tamariz, Monica. 2008. Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon* 3(2). 259–278.
- Vitevitch, Michael S., Kit Ying Chan & Steven Roodenrys. 2012. Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language* 67(1). 30–44.
- Vitevitch, Michael S. & Paul A. Luce. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9(4). 325–329.
- Vitevitch, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40. 374–408.
- Vitevitch, Michael S. & Paul A. Luce. 2005. Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language* 52(2). 193–204.
- Vitevitch, Michael S. & Mitchell S. Sommers. 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition* 31(4). 491–504.
- Warrens, Matthijs J. 2008. On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions. *Psychometrika* 73. 777–789.
- Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3). 247–274.
- Wedel, Andrew. 2007. Feedback and regularity in the lexicon. *Phonology* 24(1). 147–185.
- Wedel, Andrew. 2011. Self-organization in phonology. In Marc Van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 1–18. Hoboken: John Wiley & Sons.
- Wedel, Andrew. 2012. Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4(4). 319–355.
- Wedel, Andrew & Ibrahim Fatkullin. 2017. Category competition as a driver of category contrast. *Journal of Language Evolution* 2(1). 77–93.
- Yule, George Udny. 1912. On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society* 75(6). 579–652.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cog-2021-0087>).