Phonology (2025), **Vol. 42**:e15 1–32
doi:10.1017/S0952675725100122

PHONOLOGY

ARTICLE

An empirical study of vowel reduction and preservation in British English

Quentin Dabouis[1] and Jean-Michel Fournier[2]

[1]Laboratoire de Recherche sur le Langage (UR 999), Université Clermont Auvergne, Clermont-Ferrand, France
[2]Laboratoire Ligérien de Linguistique (UMR 7270), Université de Tours, Tours, France

Corresponding author: Jean-Michel Fournier; Email: jean-michel.fournier@univ-tours.fr

Received: 29 July 2022; **Revised:** 11 July 2023; **Accepted:** 14 July 2024

Keywords: corpus; vowel reduction; English; stress; preservation

Abstract

This article presents a dictionary-based study of vowel reduction and preservation in British English in initial pretonic position and intertonic position. The different variables which have been claimed to influence those processes are tested on a data set of over 4,500 words using regression analyses. Our results confirm the significant effects of syllable structure, position of the vowel, word frequency and opaque prefixation. They also provide weak evidence for other factors such as vowel features and the existence of a base in which the vowel bears a stress, although no clear effects of word segmentability could be found. We also report new findings, as we find that foreign words reduce less than non-foreign words; we find that [+back] vowels reduce less than [−back] vowels in initial pretonic position; and we find a difference in behaviour for vowels followed by /sC/ clusters between non-derived words and stress-shifted derivatives.

Contents

1. Introduction **2**
2. The possible determining factors of vowel reduction **4**
2.1. Syllable structure and the nature of the coda . . . 4
2.2. Position . . . 5
2.3. Morphology . . . 5
2.4. Lexical factors . . . 6
2.5. Vowel features . . . 7
2.6. Spelling . . . 8
3. Methodology **8**
3.1. Data . . . 8
3.2. Coding . . . 10
3.3. Modelling procedure . . . 13
4. Results **14**
4.1. Non-derived words . . . 14
4.2. Prefixed words . . . 17
4.3. Stress-shifted suffixal derivatives . . . 19
4.4. Non-derived words *vs.* stress-shifted derivatives . . . 23
5. Discussion **24**
5.1. The determining factors of vowel reduction . . . 25
5.2. Preservation and segmentability . . . 25

© The Author(s), 2025. Published by Cambridge University Press This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

5.3. /sC/ clusters 27
5.4. Opaque morphology and phonology 27
6. Conclusion **29**

1. Introduction

In Standard Southern British English (SSBE; a modern version of Received Pronunciation), we can observe alternations among morphologically related forms such as those in (1). We follow Bauer *et al.* (2013) in using the symbol ⬩ to mark morphological relatedness.[1]

(1) *atom* /ˈatəm/ ⬩ *atomic* /əˈtɒmɪk/
fatal /ˈfeɪtəl/ ⬩ *fatality* /fəˈtalɪti/
eponym /ˈɛpənɪm/ ⬩ *eponymous* /ɪˈpɒnɪməs/

What can be seen from the examples in (1) is that the first two vowels have different realisations depending on the position of primary stress in the word. The vowels that can bear the main stress are called 'full' vowels or 'strong' vowels, while those that cannot are called 'reduced' or 'weak' vowels.

The set of full vowels can be organised as in (2):

(2)
a.			b.			c.		
	/ɪ/	KIT		/iː/	FLEECE		/ɪə/	NEAR
	/ɛ/	DRESS		/eɪ/	FACE		/ɛː/	SQUARE
	/a/	TRAP		/aɪ/	PRICE		/ɜː/	NURSE
	/ʌ/	STRUT		/uː/	GOOSE		/ʊə/	CURE
	/ʊ/	FOOT		/əʊ/	GOAT		/ɔː/	NORTH
	/ɒ/	LOT		/aʊ/	MOUTH		/ɑː/	START
				/ɔɪ/	CHOICE			

The vowels in (2a) are usually analysed as phonologically short and are restricted to preconsonantal positions (Durand 2005). The second series, in (2b), are best analysed as closing diphthongs, even though two of them, /iː/ and /uː/, are not transcribed as such (but they regularly have diphthongal realisations such as [ɪi] and [ʊʉ] in SSBE). They may occur in all positions, including prevocalically, although there are restrictions on syllable size which may limit their appearance in closed syllables (see Harris 1994: 69; Harris & Gussmann 1998). They may undergo smoothing before schwa, are particularly affected by pre-fortis clipping, and cannot be followed by linking /r/ (Lindsey 2019: 24). Finally, the vowels in (2c) are mostly those that have emerged from the loss of post-vocalic /r/ (although some have merged with vowels that were never followed by /r/; e.g., *thought* /ˈθɔːt/, *palm* /ɑː/, *idea* /aɪˈdɪə/). They do not occur before tautomorphemic vowels, never undergo smoothing and may trigger linking /r/ before a heteromorphemic vowel (Lindsey 2019: 50). Categories (2b) and (2c) are usually analysed as phonologically long. There is some controversy regarding whether the vowels in (2) should always be analysed as having some degree of stress, which partly stems from different research and transcription traditions (see Dabouis 2020; Durand & Yamada 2023 for recent overviews).

Vowels that do not bear primary stress may be assumed to bear some level of stress (secondary or tertiary) or none at all. The main disagreement has to do with the analysis of full vowels in the positions shown in (3).

(3) a. *The syllable immediately precedes the syllable with primary stress*
Initial pretonic: /da͟ɪˈnamɪk/ *dynamic*
Intertonic: /ˌriːla͟kˈseɪʃən/ *relaxation*

[1] All the transcriptions given in this article are phonemic transcriptions adapted from Wells (2008). Adaptations include some of the main recent changes that SSBE has undergone, and which are not reflected in Wells's pronunciation dictionary but are used, among others, in Cruttenden (2014), Lindsey (2019), and Upton & Kretzschmar (2017). The symbol [æ] for the TRAP lexical set is replaced with [a]; the symbol [e] for the DRESS lexical set is replaced with [ɛ]; and the symbol [eə] is replaced with [ɛː]. Syllable boundaries (marked with spaces) are also not taken over from the dictionary.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

b. *Post-tonic vowels*
Immediately follows main stress: /ˈleɪtɛ̲ks/ *latex*
Further from main stress: /ˈdɛmənstre̲ɪt/ *demonstrate*

In the British tradition, the underlined vowels in (3) are usually analysed as unstressed full vowels, but in the American tradition, they are usually analysed as carrying some level of subsidiary stress (usually secondary or tertiary). In the rest of this article, primary stresses will be marked on orthographic transcriptions using an acute accent, and secondary stresses will only be marked in the case of full-vowelled first syllables in words with primary stress on the third syllable (e.g., *còriánder*, *kàngaróo*).

It is more difficult to define a non-controversial set of reduced vowels, as one extreme position is to assume that there are none. Indeed, Szigetvári (2017, 2018) assumes that the vowel found in words such as *cut*, *love* or *up*, which is usually represented as /ʌ/, is a 'stressed schwa'. He also claims that no vowel cannot be stressed, and that a subset may be stressed or unstressed: non-low vowels /ɪ/, /ə/ and /ʊ/ and the corresponding diphthongs /iː/, /əʊ/ and /uː/.[2] On the other hand, the two dictionary sources we use in our study, Jones (2006) and Wells (2008), posit two vowels that may be either strong or weak, /ɪ/ and /ʊ/, and three that are systematically weak, /ə/, /i/ and /u/. These latter two symbols were initially intended to represent possible variation (and thus neutralisation) between /ɪ/ and /iː/ and between /ʊ/ and /uː/. Indeed, there is no contrast between these pairs of vowels in unstressed syllables word-finally or prevocalically (e.g., *happy*, *radi̲ation*, *du̲ality*; Wells 2008; Roach 2009). Szigetvári (2017, 2022) assumes that what those dictionaries transcribe as /i/ and /u/ are unstressed realisations of /iː/ and /uː/; that /ʌ/ represents a stressed version of /ə/; and that /ɪ/ and /ʊ/ may be either stressed or unstressed. Thus, it appears that dictionaries simply use different symbols to represent what Szigetvári assumes to be vowels with different levels of stress in the case of /ə/, /i/ and /u/. We will thus call those three 'reduced' vowels, but one may assume that they are simply unstressed vowels (if one assumes that those are the only possible unstressed vowels and that cases such as those in (3) are stressed). As /ɪ/ and /ʊ/ may stand for full or reduced vowels in dictionary transcriptions, they will be excluded in all cases but one. Orthographic ⟨e⟩ is usually realised as /ɛ/, /iː/, /ɜː/ or /ɪə/, and those realisations alternate with /ɪ/ in morphologically related words (e.g., *d*/iː/*mon* • *d*/ɪ ~ ə/*moniac*; *h*/ɛ/*retic* • *h*/ə ~ ɪ ~ ɛ/*retical*). To the extent that the realisation of ⟨e⟩ as /ɪ/ is systematically given as an alternative to /ə/ and that it is almost never found in syllables carrying primary stress (except in *pretty, English, England*), it will be analysed as reduced.

There are also different views regarding what vowel reduction is. All the proposals are theory-dependent, yet all involve the loss of subsegmental material: loss of melodic features (see Giegerich 1999: §5.1.3 for a review of different post-SPE analyses using features); loss of positions, elements or heads in Element Theory (Harris 1994, 2005; Durand 2005; Backley 2011); or loss of structure in Government Phonology 2.0 (Pöchtrager 2022). There is also controversy regarding the nature of the process, if it is to be analysed as a process at all. Chomsky & Halle (1968) derived most schwas by rule from underlying full vowels, but most later analyses have rejected 'free-ride' derivations for non-alternating schwas, and analyses in Optimality Theory have captured reduction behaviour in pretonic environments mainly through *Clash constraints (see, e.g., Pater 2000), and posit surface forms as underlying when there is no related form with a different vowel. However, Szigetvári (2020: 165) claims that 'vowel reduction is not a phonological rule of present-day English, it is a historical relic' (see also Szigetvári 2018). He argues that it is a purely lexical matter which has nothing to do with phonology. While we would agree on the lexical character of reduction, to the extent that it is partly unpredictable, it is, as we will see, quite predictable in certain environments, and the generalisations that govern the distribution of full and reduced vowels are therefore surely part of the linguistic knowledge of English speakers. If one assumes it to be a phonological process, it must probably be analysed as a form of lexical redundancy rule (Jackendoff 1975) – which is the kind of generalisation found at the stem level of stratal models such as Stratal Phonology (Bermúdez-Otero 2012, 2018) – mainly because it sustains

[2] In his analysis, diphthongs are actually vowel–glide sequences, and so /iː/, /əʊ/ and /uː/ are analysed as /ɪj/, /əw/ and /ʉw/, thus in clear correspondence with the short monophthongs /ɪ/, /ə/ and /ʉ/ (the last being his transcription for the vowel in *foot*).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

a number of lexical exceptions. As for Szigetvári's (2018: 86) argument that 'it is mainly the extreme conservatism of English spelling' that suggests derivations such as /a/ → /ə/ (as in *atom* ⬩ *atomic*), we would argue that, English having long been a written language, and English speakers often being literate, it is quite possible that the phonological system of English functions differently from those of languages without such longstanding and widespread literacy. There is, in fact, considerable evidence that the phonological system is strongly restructured in the process of learning orthography, and so it is possible that what some would deem 'pure' phonology is quite restricted in English (see Dabouis 2023a). Therefore, we argue that generalisations about vowel reduction must be part of the linguistic abilities of speakers, either as lexical redundancy rules or as graphophonological rules.

Let us now get to the aim of this article. Our aim is not to argue for one analysis over another regarding the stressed or unstressed status of the vowels in (3), nor the exact nature of vowel reduction. Our aim is to study the generalisations that have been made in the literature regarding the kinds of vowels found in words such as those in (3a). In those environments, it is common to find /ə/, as in the examples in (1), but it is not systematic, as illustrated by the examples in (3a). As will be seen in the next section, numerous generalisations can be found in the literature regarding 'vowel reduction' or 'destressing', which we analyse as essentially the same process. However, no attempt has been made to put them all to the test using data in a multifactorial analysis and establish whether they all hold. Moreover, we will test factors which have only been hypothesised to have an effect, such as foreignness. We are also particularly interested in the behaviour of vowels in the environments shown in (3a), that is, intertonic position (e.g., *ànacónda*, *còndensátion*, *òbsoléte*, *ùnanímity*) and initial pretonic position (e.g., *alúmnus*, *doméstic*, *harmónic*, *provérbial*). Those positions are particularly interesting because it has been argued that the existence of a morphologically related word (usually, the local base) reduces the chances of vowel reduction. This is mostly possible in those two positions, and it can be illustrated by the famous contrast between *còn*d/ɛ~ə/*nsátion,* which is derived from *cond*/é/*nse* and may have an unreduced second vowel, and *còmp*/ə/*nsátion,* which is derived from *cómp*/ə/*nsate* and can only have a reduced second vowel (Chomsky & Halle 1968: 112–126). That observation was central to the introduction of the transformational cycle by Chomsky & Halle, which is a key feature of generative models.

The article is structured as follows. We begin by reviewing all the possible determining factors of vowel reduction which have been proposed in the literature (§2). Then, we detail the methodology of the study (§3) before we present the results (§4) and discuss their implications (§5).

2. The possible determining factors of vowel reduction

We will only focus on the claims which have been made about words that are neither compounds (e.g., *airplane, blackboard, greenhouse,* but also neoclassical compounds such as *cardiovascular, heterosexual, psychology*) nor semantically transparent prefixed words (e.g., *co-author, deconstruct, remigration*),[3] and will mainly deal with claims that apply to the two positions studied in this article (initial pretonic and intertonic).

2.1. Syllable structure and the nature of the coda

It has been claimed that vowels in closed syllables are less likely to reduce than vowels in open syllables (e.g., *fantástic*, *èxaltátion vs. horízon*, *dèprivátion*; Burzio 1994: 113; Fudge 1984; Halle &

[3]Only semantically transparent prefixed words will be studied in this article. They are quite generally assumed to be almost systematically non-reduced ('stressed' in certain analyses; Fournier 2010: chapter 1; Siegel 1974: 136–139; Raffelsiefen 1999), which goes hand in hand with the absence of reduction. Such prefixes have been assumed to be phonologically independent from their base (e.g., in Prosodic Phonology they are assumed to constitute their own phonological word; Booij & Rubach 1984). Therefore, it can be assumed that, if they are attested in pronunciation dictionaries without stress marks, we should expect them to maintain full vowels.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Keyser 1971). This is sometimes expressed differently by saying that initial pretonic light syllables undergo vowel reduction (or 'destressing'; Halle & Vergnaud 1987: 239; Hayes 1982; Selkirk 1980, 1984: 119). Moreover, it has been claimed that in closed syllables, the nature of the coda may also impact vowel reduction, with claims that vowels in syllables closed by obstruents are less likely to reduce than vowels in syllables closed by sonorants (e.g., *À͟lexánder vs. gòrgo͟nzóla*; Pater 2000) and that vowels in syllables closed by non-coronals are less likely to reduce than vowels in syllables closed by coronals (Ross 1972; Fudge 1984; Burzio 1994, 2007; Dahak 2011). Among open syllables, it has been claimed that vowels followed by another vowel are less likely to reduce (Chomsky & Halle 1968: 111; Deschamps *et al.* 2004: 217; Dahak 2011).[4]

It has also been claimed that vowel reduction may be influenced by weight interactions between syllables, as vowels found in syllables closed by non-coronal obstruents should not reduce if they are preceded by a heavy syllable, whereas vowels closed by any type of consonant should reduce if the preceding syllable is light (Fidelholtz 1966; Ross 1972; Hayes 1982; Pater 1995, 2000). This is often called the 'Arab Rule', in reference to two North American idiolectal pronunciations of the word *Arab*, /ˈærəb/ and /ˈɛɪræb/. Although these claims were made for American English, a recent study has confirmed empirically that the phenomenon exists in British English as well (Dabouis *et al.* 2020). However, the study focused on disyllabic words ending in a non-coronal obstruent and cannot confirm that this phenomenon extends to longer words or that the weight interaction observed is absent with coronal final consonants.

2.2. Position

Another important factor is the position of the vowel in the word and especially relative to stresses (e.g., whether it is pretonic or post-tonic). Many have reported a specificity of the initial pretonic position, in which reduction is far less common than in other positions (Chomsky & Halle 1968; Halle 1973; Liberman & Prince 1977; Selkirk 1984; Deschamps 1994; Deschamps *et al.* 2004). This may be attributed to the inherent strength of the initial position, and this has been analysed in Strict CV by Dabouis *et al.* (2020) as an effect of the empty CV unit posited by Lowenstamm (1999) at the left edge of the word. Positions have been noted to be relevant in post-tonic contexts as well. Dahak (2011) reports different rates of vowel reduction in different post-tonic positions depending on the number of syllables of the word and the position relative to the primary stressed syllable.

2.3. Morphology

There have also been claims on the influence of morphology on vowel reduction. The first was Chomsky & Halle's (1968: 112– 126) claim on the *condensation–compensation* contrast, which we discussed in §1. They claim that in intertonic position, having a base with stress on the second syllable (e.g., *condénse*) can protect the vowel of that syllable against reduction in the derivative in which primary stress is on its third syllable. More recently, it has been argued that the frequencies of the derivative and the base are relevant. In a study of *-ation* derivatives in which the second syllable is closed by a sonorant and which have a base with second-syllable stress (e.g., *condémn* ⬩ *condemnation*), Hammond (2003) finds that vowel reduction in the second syllable is more likely if the base frequency is higher and if the derivative frequency is higher. However, Collie (2007: 182–186) replicated Hammond's regression analysis and found that result to be statistically unreliable. Others, following Hay (2001, 2003), have argued that the relevant frequency measure is the relative frequency of the base and of the derivative. In Hay's model, the 'segmentability' of a word is related to lexical storage and how complex words are accessed in the mental lexicon, as more segmentable words are assumed to be accessed in long-term memory through their constituents, while less segmentable words are accessed directly as whole

[4] The verification of this claim will be strongly influenced by the position adopted regarding the analysis of what Wells (2008) transcribes as /i/ and /u/ (see discussion in §1).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

forms. Relative frequency has therefore been described as an indirect measure of the segmentability of a complex word (Plag & Ben Hedia 2018). As it is indirect, it is also possibly imperfect, and so it can be complemented by additional measures of segmentability such as semantic transparency. Segmentability is predicted to have phonological effects, so that if the base is more frequent than the derived word, the derived word is more likely to preserve phonological properties from its base, but if the base is less frequent than the derivative, then the latter is less likely to preserve such properties. Thus, segmentability may interact with absolute frequency of the derivative, which has also been claimed to affect vowel reduction, as will be seen in §2.4. In that model, vowel reduction is expected to be less likely if the base is more frequent than the derivative, as shown in (4), which are examples taken from Bermúdez-Otero (2012), based on observations made by Kraska-Szlenk (2007: §8.1.2). However, both Hammond (2003) and Kraska-Szlenk (2007) use very small samples and do not test frequency effects in other environments.

(4) *Frequencies (tokens per 10^6 words in the spoken section of the Corpus of Contemporary American English) of some base–derivative pairs* (Bermúdez-Otero 2012: 37)

			Base		Derivative
a.	*Cyclic stress*				
	cond[ɛ́]mn	còmd[ɛ̀]mn-átion	7.09	>	2.57
	imp[ɔ́]rt	ìmp[ɔ̀]rt-átion	5.15	>	0.62
b.	*Variable stress*				
	cond[ɛ́]nse	còmd[ɛ̀ ~ ə]ns-átion	0.28	≈	0.22
c.	*Non-cyclic stress*				
	cons[ɜ́]rve	còns[ə]rv-átion	1.65	<	9.11
	tránsp[ɔ́]rt	tránsp[ə]rt-átion	7.23	<	23.54

Other relative frequency effects have been reported for processes such as second-syllable preservation failure (e.g., *antícipate* ⬩ *antìcipátion* ~ *ànticipátion*; Collie 2007, 2008), exceptional second-syllable stress preservation (e.g., *adópt* ⬩ *adòptée*; Dabouis 2019) or morphological gemination (e.g., *i*[ɹː]*ational*; Dabouis *et al.* 2023), and these different results have been interpreted using Hay's model. Therefore, these different claims mean that such morphological effects need to be controlled for in relevant positions, including the initial pretonic position, which has not been investigated in that regard in the literature.

The second type of influence that morphology has been claimed to have on vowel reduction concerns semantically opaque prefixed words such as *contain, deceive, reduce, suspect.* Most major works on English phonology have observed that the vowel of their prefix reduces systematically in initial pretonic position, even if it is in a closed syllable (Chomsky & Halle 1968: 118; Halle & Keyser 1971: 37; Liberman & Prince 1977: 284–285; Guierre 1979: 253; Selkirk 1980; Hayes 1982; Halle & Vergnaud 1987: 239; Pater 2000; Hammond 2003; Collie 2007: 129, 215, 318–319).[5] Note that this may be true of SSBE and General American but not of northern varieties of England, where a number of prefixes in closed syllables maintain full vowels (e.g., *conclude* /kɒŋˈkluːd/, *advance* /adˈvans/, *substantial* /sʌbˈstanʃəl/; see Cruttenden 2014: 139).

2.4. Lexical factors

Reduction has also been claimed to be influenced by lexical factors. One of them is frequency, as Fidelholtz (1975) claims that more frequent words are more likely to undergo reduction than less frequent ones, which has been confirmed by later studies (Bell *et al.* 2009; Clopper & Turnbull 2018).

[5] Another observation made in the literature regarding opaque prefixed words and vowel reduction is that disyllabic nouns with initial stress that have a related verb with final stress (e.g., *escort, import, project*) do not show reduction of their second vowel (Fudge 1984: 32, 167; Poldauf 1984: 38). However, in her large-scale dictionary-based study, Dahak (2011: 235) reports that this behaviour extends to words without a morphologically related word with final stress and that, unlike in non-prefixed words, vowel reduction is quite rare in disyllabic opaque prefixed words with initial stress (e.g., *abject, adverb, expert, oblong*).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

The frequency of units smaller than the word has also been proposed to be a factor in vowel reduction in opaque prefixed words. Hammond (2003), who bases his analysis on Fidelholtz's work, claims that lexical items with a high frequency tend to reduce, and that the high frequency of Latinate prefixes would explain why they are often reduced (see §2.3). However, this seems to interact with word frequency, as Pater (2000) notes that the first vowels in common prefixed words (e.g., *admire, compose, embrace, protect*) tend to reduce more than in less common prefixed words (e.g., *adsorb, exogamy, obtund, protrude*). The frequencies of the words given by Pater were collected by Cho (2004) and, for these words at least, the generalisation holds, as words whose prefix does not reduce never have a high frequency.

Another lexical factor that has been put forward for phonetic reduction is neighbourhood density, which is 'typically defined as the number of words that differ from a target word by one phoneme insertion, deletion, or substitution' (Clopper & Turnbull 2018: 30). Words with more neighbours may be more difficult to identify and so are less likely to undergo reduction.[6]

Finally, Dabouis & Fournier (2022) analyse the English lexicon as divided into several subsystems which have different phonological, morphological, graphophonological and semantic properties, and which are defined by the perceived foreignness or 'learnedness' of words. They suggest that foreign words (in their system, those belonging to the lexical subsystems §French and §Foreign) might be less prone to vowel reduction than native vocabulary, although this effect should be teased apart from that of word frequency. Dahak (2009) also mentions this factor and reports that half of the 34 words that she tagged as 'borrowed' in her study on vowel reduction in intertonic position have a full vowel. She suggests that 'the level of integration of a word into the system' can be a determining factor.

2.5. Vowel features

We are aware of only two proposals regarding how certain vowel features may affect vowel reduction. Chomsky & Halle (1968) have two such claims: tense vowels do not reduce prevocalically, and only unstressed low vowels reduce in final position. However, proposals such as these are difficult to test as they depend on what one assumes to be the underlying vowel from which the surface reduced vowel would be derived (in a generative model such as Chomsky & Halle's, where surface [ə] is often derived from underlying full vowels). In monomorphemic words, we have no way to determine what the underlier is if the surface vowel is [ə]. However, in stress-shifted derivatives, or in words which have a cognate in which the vowel under scrutiny is full, that vowel may be assumed to be the underlier. In order to broaden the potential scope of these claims, we will test whether certain vowel classes behave differently from others with regard to vowel reduction.

One way to go around this issue could be to use spelling, which is systematically available regardless of the surface value of the vowel. This is the approach adopted by Deschamps (1994: 111–112), who notes that ⟨o⟩ and ⟨u⟩ do not reduce word-finally, unlike other vowels, or by Tokar (2019), who reports different rates of vowel reduction for ⟨a⟩ and ⟨o⟩ in initial pretonic position in open syllables (93% *vs.* 69%). However, if one assumes that orthography represents underlying vowels, it is unclear what subsegmental properties should be assumed for these vowels. For example, in British English, the four most common realisations of ⟨a⟩, namely /a, eɪ, ɑː, ɛː/, vary considerably in height. Among vowel monographs, we could identify two series that almost systematically differ in potential backness: ⟨a, e, i, y⟩ *vs.* ⟨o, u⟩.

[6] There are other factors which would be relevant for phonetic reduction in speech, as discussed by Clopper & Turnbull (2018), who say that reduction is influenced by how difficult the processing of words is for both speakers and listeners, and that this difficulty is influenced by factors such as semantic predictability, discourse mention and speaking style, along with lexical factors such as frequency and neighbourhood density. Bell *et al.* (2009) also report that the predictability of content and function words (along with their frequency) impacts their duration.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

2.6. Spelling

Spelling is not often included among the possible determining factors of phonological processes, but some authors have done so, claiming that vowels spelled with digraphs reduce less than vowels spelled with monographs, particularly in initial and final position (e.g., *augmént, Eurásian vs. pathétic, forénsic*; Dahak 2011; Deschamps *et al.* 2004: 217). In response to earlier presentations of this work, it has been suggested to us that this parameter might overlap with that of individual vowels, as digraphs almost systematically represent long vowels, and certain vowels (/ɔɪ, aʊ/) are only represented with digraphs, so the potential effect of spelling would have to be teased apart from that of vowel features such as length or backness.

3. Methodology

3.1. Data

The use of large data sets has long been the exception rather than the rule, and, as noted by McMahon (2001: 424), '[T]here is undoubtedly a problem in phonology, especially the sort that rather distances itself from phonetics, of reliance on stock examples and introspection'. This problem surely has affected studies on vowel reduction, as we are aware of only a handful of large-scale empirical studies on the issue, each of which is limited in scope: Hammond's (2003) study of words of the *condensation* type (which provides no information about how many words were investigated and how they might differ from non-derived words); the large study of post-tonic vowels in Dahak (2011), which has not sought to investigate the interactions between all possible factors, so that some of the results that Dahak attributes to one factor may in fact be caused by another; Tokar's (2019) study of orthographic ⟨o⟩ in initial pretonic position; the study of the 'Arab Rule' in Dabouis *et al.* (2020); and Zhang's (2021) study of vowel reduction in deverbal nouns in *-(at)ion*, which takes only a limited number of factors into consideration. Therefore, using pronunciation dictionary data seems a good starting point, as they have the advantage of giving access to large numbers of words, with several pronunciations listed and a relative uniformity of the idiolect that is represented (even though it is to some extent an artificial idiolect).[7] They also present limitations such as questionable syllabification choices (Ballier & Martin 2010), the hybrid nature of the transcription used (Dahak 2006) or the unavoidable presence of errors. Because of these drawbacks, the present study should be later complemented with studies using other kinds of data such as natural speech data or judgement tasks.

We chose to focus on two positions:

Initial pretonic: The first syllable of a word, immediately followed by a stressed syllable (e.g., *arríve, dextérity, herétical*).[8]

Intertonic: The second syllable of a word, immediately followed by the syllable with primary stress and preceded by a syllable with secondary stress (e.g., *rèlaxátion*). Words listed as having a variant in which the first syllable is unstressed and the second syllable carries secondary stress

[7] To get a sense of the reliability of the dictionary data, the first author compared the pronunciations given in the dictionaries used in this study with oral data in Youglish (https://youglish.com/, consulted between 12 January and 9 March 2021), in different syllabic configurations, with different orthographic vowels, for both monomorphemic and stress-shifted derivatives and only in initial pretonic position. Two hundred words were considered, and a maximum of 10 tokens per word were listened to, for a total of 1,316 tokens. Vowels were classified by ear as full or reduced. The main result is that words whose main pronunciation is given as reduced by the dictionary are indeed reduced in 90% of cases, while those given as full are full in 70% of cases, with more reduced vowels in open syllables (76%) than in closed syllables (38%). This suggests that the dictionary data are quite reliable overall; that the main pronunciation is the most representative of what speakers produce; and that there is (unsurprisingly) more variability in oral data than in the dictionary data.

[8] In the case of words whose second syllable is marked in the dictionary as having secondary stress and where the first syllable contains a full vowel, some might argue that both have secondary stress. However, note that such cases are almost exclusively attested among prefixed words (36 prefixed words and 2 non-prefixed) and that they are not numerous enough to have any significant impact on the results presented here.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

(e.g., *depàrtméntal*) are left out, so as to keep only words in which the first syllable is stronger than the second syllable (but see Dabouis 2019 for a study of that pattern).

As the only effects of morphology that we seek to study are the two detailed in §2.3, namely possible vowel preservation from a base in which the vowel has main prominence, and the effects of prefixation, we will restrict our investigation to three morphological categories:

1. Monomorphemic words (e.g., *acacia, cadastre, elite, macabre, tarantula*) and words formed of a bound root and a suffix (e.g., *ambition, hermetic, sporadic*), which are used as a reference point for how 'vowel reduction' functions in the absence of any morphological influence. Those two categories are treated together because all cyclic models assume them to be computed in a single pass through the phonology (as roots do not trigger phonological computation), and so they should behave in the same way. From now on, these words will be referred to as NON-DERIVED words;
2. Prefixed words with a monosyllabic prefix, using a broad definition of prefixation including historically prefixed words which are synchronically semantically opaque (e.g., *accede, believe, collect, defend, elapse, extent, offend, persist, presume, promote, recite, succeed*) in order to test the claims made in the literature (see §2.3). We will distinguish such opaque prefixed words, in which the prefix (and sometimes the root, which may be bound) contributes no clear meaning, from transparent prefixed words, in which both constituents have clearly identifiable meanings (e.g., *asymmetry, co-author, decentralize, preamplifier, reactivate, subarctic, transnational, unaltered, unwrap*). Certain words could not be treated as straightforwardly opaque or transparent, as there is a semantic contribution of the prefix to the semantics of the whole word, but the meaning of the whole word is not compositional or the root is bound. Semantic transparency is gradient, and so we sought to study the constructions which are quite clearly opaque or quite clearly transparent. Therefore, we excluded 289 words in the 'grey area' between the two (e.g., *cohabit, degenerate, empower, extract, resuscitate, transform*). This morphological category will be included only in our consideration of initial pretonic position;
3. Stress-shifted derivatives (e.g., *vítal* • *vitálity*; *ínfərm* • *ìnformátion*), which provide useful information regarding how different vowels (or subsegmental features) impact reduction and whether the existence of a base with main prominence on the vowel under scrutiny affects reduction and, if so, how. We included words in those data sets only if a base could be identified and if that base has a frequency above zero in SUBTLEX-UK (van Heuven *et al.* 2014).

Other morphological categories such as compounds, neoclassical compounds or suffixed words without stress shift were not considered. For the reasons exposed in §1, we did not keep any occurrences of /ʊ/, nor of /ɪ/ spelled ⟨i⟩.

Two different sources have been used for the two positions and, in both cases, only British pronunciations are considered. For initial pretonic position, we automatically extracted all the words with no stress mark on their first syllable in Jones (2006), and manually identified the relevant morphological categories. Jones (2006) is a pronunciation dictionary which has around 80,000 entries, with both British and American pronunciations. For intertonic position, the data were taken from Wells (2008), but were extracted from the data sets used in Dabouis (2016), which are available online (at https://halshs.archives-ouvertes.fr/tel-01414997/file/Annexes.pdf): monomorphemic words, bound roots plus a suffix, and derived words with a base with main prominence on its second syllable.

Wells (2008) is also a pronunciation dictionary, which has around 83,000 entries and also lists British and American pronunciations. Both dictionaries use phonemic transcriptions (although see §1 on their use of /i/ and /u/), and may list several possible pronunciations for a given word. For example, the entry in Wells (2008) for the verb *extract* reads '**ɪk** ˈ**strækt** ek-, ək-'. The main pronunciation is shown in bold, and the other two possible pronunciations for the first syllable are variants. Following Hammond (2003), the transcriptions were converted into a four-point scale (1: Full; 2: Full ~ Reduced; 3: Reduced ~ Full; 4: Reduced). If only a full pronunciation is given, the vowel was coded as 1; if only a reduced pronunciation is given, the vowel was coded as 4; and if both a full and a reduced pronunciation

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Table 1. *Word counts in the different data sets of the study.*

	Initial pretonic	Intertonic
1. Non-derived	1,234	474
2. Prefixed	1,997	—
3. Stress-shifted derivatives	590	199

are given, vowels were coded as 2 or 3 depending on the order in which the variants are listed. There is one case for which possible variation is represented within the main pronunciation, /ə(ʊ)/, and so we treated it as Full ~ Reduced (coded as 2).[9]

The reason that we used two different sources is that the data set for initial pretonic position was initially designed as a follow-up to a study on the 'Arab Rule' (Dabouis *et al.* 2020), and we later expanded the investigation to intertonic position as we had an already available cleaned-up data set in Dabouis (2016). In both cases, all the relevant cases are included, as long as they fit the morphological categories described above. This means that other morphological categories (e.g., compounds, neoclassical compounds, words with neutral suffixes) are not included. In intertonic position, we did not keep words whose second syllable is part of a historical prefix (e.g., *recollect, supersede*), as these prefixes have been claimed to reduce systematically (see §2.3; this was indeed true in all but one case), and they were not present in sufficient numbers to constitute a separate category for statistical testing. In intertonic position, derivatives containing a semantically transparent prefix (e.g., *amoral* • *amorality*) were not kept in the data set, as the intertonic vowel is systematically identical to that found in the corresponding non-prefixed word. We also excluded the few cases in which the vowel under consideration is spelled differently in the base and in the derivative (e.g., *rev**eal*** • *rev**e**lation*) as spelling was a variable to be tested and in these cases, it would not have been clear which spelling to consider. Note that some of the words used have variation in the position of stresses.

In order to gauge the extent to which the two sources agree regarding vowel reduction, we extracted a random sample of 100 entries in each of the three inventories of initial pretonic position (which are taken from Jones 2006), amounting to 300 words in total. Then we collected the pronunciations given in Wells (2008) and checked whether the two dictionaries agreed in giving a full or reduced pronunciation of the relevant vowel. There are 25 words which are not listed in Wells, so the comparison can be done only on 275 words. We found that the two dictionaries give strictly identical information in 226 cases (82%), and that they agree on the main pronunciation in 254 cases (92%). The differences mainly have to do with the fact that Wells sometimes uses /u/ where Jones uses /uː/ (12 different entries). The remaining cases include five entries for which the two dictionaries give the same pronunciations but reverse the order of the variants, and four are rather rare and foreign words (*batik, pesewa, razoo, sapele*). Thus, the two dictionaries largely converge on the data they provide regarding vowel reduction and can be assumed to be comparable. However, because there are some differences, we will never collapse the two data sets into a single one in the following analyses; the two positions will always be analysed separately.

The number of words for each data set is shown in Table 1.

3.2. Coding

We coded the data based on the different variables proposed in the literature which we reviewed in §2. The variables used and how they were coded are detailed below.

SyllableStructure: The literature discussed in §2.1 argues that there is a difference between open and closed syllables. We coded this variable following standard syllabification procedures, maximising

[9]The fully annotated data and R scripts are available on OSF: https://osf.io/qbcnv/.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

onsets while respecting the sonority contour. This means that sequences with rising sonority that can constitute a well-formed onset (e.g., /br/, /pl/, /lj/) were analysed as branching onsets, while those with level or falling sonority (e.g., /pt/, /kt/, /lt/, /nd/) and clusters of rising sonority which are not well-formed onsets (e.g., /fg/, /ps/, /gn/) were analysed as coda–onset sequences. Syllables with no coda are coded as OPEN (e.g., *lasagne, instrument*), and those with a coda are coded as CLOSED (e.g., *campaign, volunteer*).

However, there are two problematic structures for which we had to make choices. First, /sC/ clusters are a well-known problem for syllabification, as they are the only attested word-initial clusters with falling sonority (see, e.g., Scheer & Ségéral 2020; Goad 2012), and so their word-internal syllabification is an issue. The second problematic structures are vowels followed by historical coda /r/. As the variety of English that we are focusing on is non-rhotic, it is unclear whether we should assume that coda /r/s are still present underlyingly or not. Statistical tests were conducted in the non-derived data set to determine how to treat those problematic structures. No difference was found for reduction rates between vowels followed by ⟨rC⟩ and vowels followed by other consonant clusters which may not form branching onsets. Therefore, vowels followed by ⟨rC⟩ were coded as closed. However, vowels followed by /sC/ were found to be statistically significantly different from both open and closed syllables, and so were coded as a distinct category, sC.

We then coded two variables reflecting the place and manner of the coda, which have been claimed to affect vowel reduction (§2.1):

CODA-PLACE: Codas were coded as CORONAL (/n/, /l/, /d/, /t/, /z/, /ʃ/) or NON-CORONAL (/k/, /g/, /p/, /b/, /m/, /ŋ/).

CODA-MANNER: Codas were coded as OBSTRUENT (/p/, /t/, /k/, /b/, /d/, /g/, /z/, /ʃ/) or SONORANT (/n/, /m/, /ŋ/, /l/).

WEIGHTSI: In order to test possible interactions between the first two syllables, we coded the weight of the first syllable in the intertonic data set as HEAVY (e.g., *aviation, trampoline*) or LIGHT (e.g., *magazine, coriander*). This variable is designed to control potential weight interactions between the first two syllables which may resemble the 'Arab Rule' as discussed in §2.1. The behaviour noted in the literature is that vowel reduction may be affected by the weight of the preceding syllable (in the case of the 'Arab Rule', the place of articulation of the coda consonant in the second syllable is also relevant). Quite logically, this variable is only relevant in intertonic position, as no syllable precedes the vowel in initial pretonic position.

SPELLING: Vowels were coded as MONOGRAPH (e.g., ⟨a⟩, ⟨o⟩, ⟨u⟩) or DIGRAPH (e.g., ⟨ai⟩, ⟨au⟩, ⟨ow⟩), as we saw in §2.6 that there are claims in the literature that digraphs reduce less often than monographs.

GRAPHEME: The specific grapheme was coded. This is one way to test vowel features, as there are reports that different orthographic vowels behave differently (§2.5). As there are too few occurrences of each different digraph, statistical tests including this variable were only conducted on monographs, excluding ⟨i⟩, as most occurrences were excluded from the general data set on the grounds that ⟨i⟩ corresponding to /ɪ/ is uninterpretable in terms of vowel reduction (see §1), and ⟨y⟩ as there were too few occurrences. Therefore, there are only four possible values for this variable: ⟨a, e, o, u⟩.

LOGFREQUENCY: Token frequencies were collected from SUBTLEX-UK (van Heuven *et al.* 2014), and were log-transformed (as $ln(x + 1)$) so as to resemble the way 'humans process frequency information' (Hay & Baayen 2002: 208). This was done to test the common observation that high-frequency words tend to undergo more reduction than low-frequency items (§2.4).[10]

FOREIGN: In response to earlier presentations of this work, in which we had reported word frequency to be a significant predictor of vowel reduction, it was suggested to us that this effect could (at least partly) be attributed to the fact that many low-frequency words were foreign. The assumption is then

[10]We also tested the claim that neighbourhood density affects vowel reduction. We collected values from Shaol & Westbury (2010) for non-derived words. However, this source did not contain enough data for us to be able to test this predictor properly, and in preliminary analyses, we found no effect of that variable in the data sets in which we had the most data.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

that foreign words would be less likely to reduce than non-foreign words. Although we do not know of any published research on this issue, there are proposals that foreign words in English may behave differently from more integrated words (Pater 1994; Dabouis & Fournier 2022). Therefore, in order to tease apart the effects of frequency from those of foreignness, it was necessary to identify which words should be treated as foreign. This is not easily done, and so we relied on different formal characteristics that are available to speakers, following Dabouis & Fournier (2022). We used three characteristics in a stepwise fashion (i.e., words identified as foreign for one characteristic were not considered for the following criterion):

1. *Word endings*: Certain orthographic endings appear almost exclusively in loanwords, and we identified nine such endings in our data: ⟨-Ca⟩, ⟨-Ci⟩, ⟨-Co⟩, ⟨-Cu⟩, ⟨-eur⟩, ⟨-euse⟩, ⟨-aise⟩, ⟨-é(e)⟩ and ⟨-V(rr)h⟩. For example, here are the first ten words in ⟨-Ci⟩ in our initial pretonic data set: *acouchi, adzuki, afghani, agouti, aioli, basmati, bikini, borlotti, bouzouki, chapatti.*
2. *Foreign spelling-to-sound correspondences*: These have been established in previous studies such as Carney (1994), Deschamps (1994) and Trevian (1993). They include irregular correspondences for stressed vowels such as ⟨a⟩–/ɑː/ (as in *ban**a**l, mén**a**ge*) and ⟨i⟩–/iː/ (*el**i**te, past**i**s*); consonant correspondences such as ⟨ch⟩–/ʃ/ (***ch**andelier, moustac**he***) and ⟨g⟩–/ʒ/ (*colla**g**e, in**g**enue*); and non-silent final ⟨e⟩ (*anemon**e**, furor**e***).
3. *Semantics referring to foreign cultures*: A number of loanwords may only be identified through their meaning when it refers to objects, people or customs of the peoples whose languages were the sources of these loans. We identified mainly meanings referring to foreign currencies (e.g., *koruny, pistole, rupee*), food and drinks (*champagne, kebab, trepang*), functions (*hussar, savoy, ukase*) and objects (*palankeen, pirogue, sitar*). Details of the categories used can be found in the full data set on OSF (see https://osf.io/qbcnv/).

This variable was coded only for non-derived words. Although this methodology might miss some of the targeted words and identify as foreign certain words which may not be perceived as such (such as *banana, charisma, police, potato*), it should allow us to capture most of the effects of foreignness, if there are any. The number of words identified as foreign is quite significant: 600/1235 (49%) in the initial pretonic data set and 275/474 (58%) in the intertonic data set.

Finally, certain variables were coded only for stress-shifted derivatives:

Morphology: In initial pretonic position, the presence of a semantically opaque prefix was coded: Prefixed (e.g., *contextual, objectify, proverbial*) *vs.* NonPrefixed (*articular, musician, solidity*). Although the literature quite generally reports that opaque prefixes reduce in initial pretonic position (§2.3), there is no established uncontroversial and reproducible way to identify such prefixes. Therefore, we used the etymological information given in the online *Oxford English Dictionary* (https://www.oed.com/) to establish whether we should treat a word as prefixed. Although certain words may have lost the formal characteristics which would allow for the recognition of these prefixes, we assume that most of them have kept such characteristics (see Dabouis & Fournier 2025), which has been argued to be mainly the distributional recurrence of the prefix (e.g., *conceptual, confiscatory, constituent*) and root (e.g., *aspectual*; cf. *inspect, prospect, suspect*) and medial consonant clusters which are phonotactically illicit in simplex words (e.g., /kspl/ is only attested in words in *ex-* such as *explain, explicit, exploit*).

LogFrequency-Base: Log-transformed frequency of the base (as $ln(x + 1)$), also taken from SUBTLEX-UK.

RelativeFrequency: Ratio of the log-frequency of the derivative and the log-frequency of the base. Thus, a relative frequency of less than 1 means that the base is more frequent than the derivative, and a relative frequency greater than 1 means that the base is less frequent than the derivative.

SemanticTransparency: Derivatives for which the base appears explicitly in the definition of the derivative in a general dictionary (Dictionary.com, consulted in May 2019) were coded as Transparent. Others were coded as Opaque.

In order to test for vowel features to see if certain natural classes have specific behaviours (see §2.5), different models and analyses of English vowels were tested. We tried using Backley's (2011) Element

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Table 2. *Vowel features based on Jensen (2022).*

		BACK	HIGH	LOW	ROUND
DRESS	/ɛ/	No	No	No	No
TRAP	/a/	No	No	Yes	No
LOT	/ɒ/	Yes	No	Yes	Yes
STRUT	/ʌ/	Yes	No	No	No
FLEECE	/iː/	No	Yes	No	No
GOOSE	/uː/	Yes	Yes	No	Yes
THOUGHT/NORTH	/ɔː/	Yes	No	No	Yes
PALM/START	/ɑː/	Yes	No	Yes	No
NURSE	/ɜː/	Yes	No	No	No
FACE	/eɪ/	No	No	No	No
GOAT	/əʊ/	Yes	No	No	Yes
MOUTH	/aʊ/	No	No	Yes	No
PRICE	/aɪ/	Yes	No	Yes	No
CHOICE	/ɔɪ/	Yes	No	No	Yes
NEAR	/ɪə/	No	Yes	No	No
SQUARE	/ɛː/	No	No	No	No
CURE	/ʊə/	Yes	Yes	No	Yes

Theory model, by coding vowels as containing or not containing one of the three elements used by Backley. A second option was to use Jensen's (2022: 64–66) analysis of English vowels using binary features. That second option turned out to perform much better in statistical analyses, and so here is how it was implemented. We used four variables to represent the four features used by Jensen: BACK, HIGH, LOW and ROUND. Vowels were coded for all four variables, with two possible values, YES or NO. As Jensen is mainly focused on American English, the vowels /ɜː/, /ɪə/, /ɛː/ and /ʊə/ are not present in his analysis, and so we inferred feature specifications from those of other vowels. The coding used for our data is shown in Table 2.

As can be seen from Table 2, those features alone cannot capture all possible vowel contrasts, and so those variables were used alongside the variable VOWELQUANTITY. For this variable, vowels in the base were coded as LONG (diphthongs and long vowels) or SHORT (the four vowels /a, ɛ, ɒ, ʌ/). This was done to test the claim that long vowels reduce less than short vowels (§2.5).

Finally, the dependent variable is **VOWELREDUCTION**, coded using a four-point scale as described in the previous section.

3.3. Modelling procedure

All statistical tests were conducted in R (v. 4.1.3; R Core Team 2023) using ordinal logistic regression with VOWELREDUCTION as the dependent variable, as this variable is a scale. This was done using the `polr` function from the `MASS` package (Ripley *et al.* 2019). Models were progressively simplified step-by-step following standard procedures (e.g., Baayen 2008). We follow Engemann & Plag (2021) in assuming that, to be maintained in a model, a variable has to meet three criteria. First, its *t*-value had to be either below −2 or above 2. Second, the AIC of the model including the variable had to be at least two points lower than the model without it. Third, a likelihood ratio test comparing the model including the variable and the model without it had to have a *p*-value lower than 0.05. A variable was included in a model only if it passed all three tests, showing that its inclusion significantly improved the model.

The residuals of the final models were analysed using the `resids` function of the `sure` package (Greenwell *et al.* 2018), which uses surrogate residuals (Liu & Zhang 2018), as ordinal logistic

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

regression models cannot be analysed directly using common tools for residual analysis, and were found to have normal distributions on both tails of the distribution. As will be explained in the following sections, certain variables were only tested on subsets of the data, as not all factors are relevant for all words (e.g., coda place and manner are applicable only to vowels in closed syllables).

4. Results

4.1. Non-derived words

This section deals with monomorphemic words and words formed of a bound root and a suffix, looking at both initial pretonic position ($n = 1,234$) and intertonic position ($n = 474$). Following the procedure described in §3.3, the final models show effects of SyllableStructure, Spelling, LogFrequency, Foreign and, for intertonic position, WeightSi (recall that this variable is not relevant for initial pretonic position, where there is no preceding syllable). The regression results for those models are shown in Table 3.

As the words containing digraphs represent a small part of the data – 60 words in the initial pretonic data (5%) and 19 words in the intertonic data (4%) – we will detail the results regarding the difference between digraphs and monographs first and subsequently deal only with monographs. To illustrate the difference between digraphs and monographs, let us consider only open syllables in Figure 1, as digraphs hardly ever occur in closed syllables.

Two things can be observed in Figure 1. First, there is a clear difference between the two positions, as reduced vowels are more common in intertonic position than in initial pretonic position. Second, we can see an obvious difference between monographs and digraphs, with digraphs more often representing full vowels than monographs. Examples are shown in (5).

(5) a. *Initial pretonic*

i. *Digraphs*:	*audition*	/ɔːˈdɪʃən/
	authority	/ɔːˈθɒrəti/ ~ /əˈθɒrəti/
	moussaka	/muˈsɑːkə/
ii. *Monographs*:	*antenna*	/anˈtɛnə/
	molest	/məʊˈlɛst/ ~ /məˈlɛst/
	falafel	/fəˈlɑːfəl/

Table 3. *Ordinal logistic regressions for non-derived words in both positions.*

	Initial pretonic			Intertonic		
Predictor	Odds ratio	95% C.I.	*p*	Odds ratio	95% C.I.	*p*
1|2	175.14	73.61 – 416.68	<0.001	4.10	0.90 – 18.66	0.068
2|3	420.21	173.15 – 1,019.83	<0.001	6.91	1.52 – 31.41	0.012
3|4	870.81	354.50 – 2,139.09	<0.001	15.13	3.28 – 69.81	0.001
SyllStr-Open	26.59	18.27 – 39.63	<0.001	16.10	8.02 – 32.87	<0.001
SyllStr-sC	8.50	4.53 – 15.99	<0.001	5.19	0.99 – 42.41	0.076
LogFq	1.16	1.10 – 1.22	<0.001	1.25	1.07 – 1.48	0.006
NatV-Monograph	22.39	11.11 – 50.29	<0.001	0.51	0.26 – 0.95	0.039
Foreign-Yes	0.76	0.60 – 0.95	0.015	6.74	1.94 – 21.01	0.002
WeightSi-Light	Irrelevant			2.36	1.31 – 4.32	0.005
AIC	2,677.849			437.126		

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

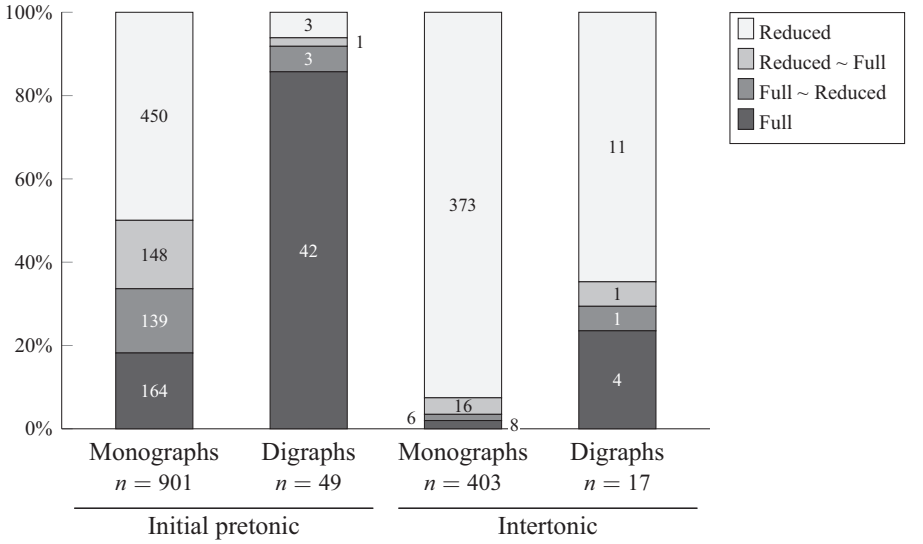


Figure 1. *Vowels found for digraphs and monographs in open syllables in non-derived words.*

b. *Intertonic*

i.	*Digraphs:*	*araucaria*	/ˌarɔːˈkɛːriə/
		demoiselle	/ˌdɛmwɑːˈzɛl/ ~ /ˌdɛmwəˈzɛl/
		connoisseur	/ˌkɒnəˈsɜː/
ii.	*Monographs:*	*incantation*	/ˌɪnkanˈteɪʃən/
		ostentation	/ˌɒstɛnˈteɪʃən/ ~ /ˌɒstənˈteɪʃən/
		pioneer	/ˌpaɪəˈnɪə/

If we now focus on monographs (n = 1, 174 and 455), the distribution of the data based on syllable structure is shown in Figure 2.

As can be seen from Figure 2, our results confirm the previous literature regarding the effects of syllable structure on vowel reduction, as there is indeed more vowel reduction in open syllables than in closed syllables. Syllables in which the vowel is followed by /sC/ seem to constitute an intermediate class, possibly because these clusters may be parsed heterosyllabically or tautosyllabically. Examples are shown in (6).

(6) a. *Initial pretonic*

i.	*Open:*	*gigantic*	/dʒaɪˈgantɪk/
		gorilla	/gəˈrɪlə/
ii.	*sC:*	*cascade*	/kasˈkeɪd/
		estate	/ɪˈsteɪt/
iii.	*Closed:*	*campaign*	/kamˈpeɪn/
		gestalt	/gəʃˈtalt/

b. *Intertonic*

i.	*Open:*	*cacoethes*	/ˌkakəʊˈiːθiːz/
		propaganda	/ˌprɒpəˈgandə/
ii.	*sC:*	*bantustan*	/ˌbantuːˈstɑːn/
		balustrade	/ˌbaləˈstreɪd/
iii.	*Closed:*	*chimpanzee*	/ˌtʃɪmpanˈziː/
		guarantee	/ˌgarənˈtiː/

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

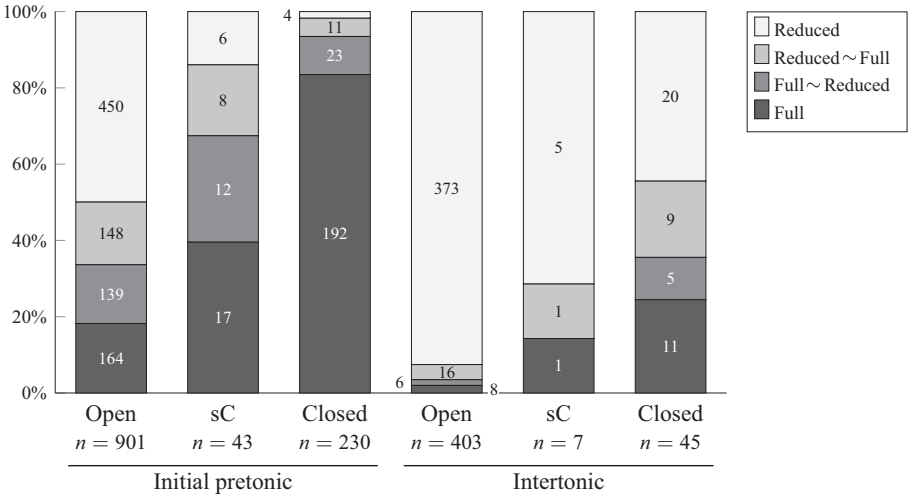


Figure 2. *Vowels found for monographs in non-derived words depending on syllable structure.*

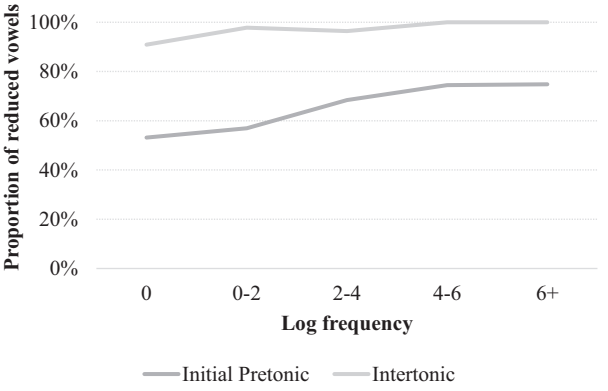


Figure 3. *Proportion of words with a reduced main pronunciation depending on their log frequency.*

Our results also confirm that high frequency implies greater rates of vowel reduction. This can be seen from Figure 3, which shows the proportion of words whose main pronunciation is reduced, for monographs in open syllables depending on their frequency. As can be clearly seen, the proportion of reduced vowels increases as frequency increases.

Our results also bring forward a new finding: foreign words undergo less vowel reduction than words that are not foreign, and this effect is independent of frequency.[11] Moreover, there is an effect of the weight of the first syllable in the intertonic position, which is reminiscent of the 'Arab Rule'. The latter could not be tested specifically, as there are too few words with a closed second syllable to test the possible difference between those closed by coronal obstruents and those closed by non-coronal

[11] There is no statistically significant difference between the frequencies of the words coded as foreign and that of those coded as non-foreign (Mann–Whitney U test; initial pretonic: $W = 164,217$, $p = 0.1885$; intertonic: $W = 25,911$, $p = 0.70$). Following a suggestion from one of the editors, we also tested for an interaction between frequency and foreignness, which was non-significant in both positions.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Table 4. *Ordinal logistic regression for the subset of words containing one of the four monographs ⟨a, e, o, u⟩.*

	Initial pretonic			Intertonic		
Predictor	Odds ratio	95% C.I.	*p*	Odds ratio	95% C.I.	*p*
1|2	6.61	4.32 – 10.12	<0.001	0.59	0.23 – 1.50	0.265
2|3	22.62	14.29 – 35.81	<0.001	1.08	0.43 – 2.71	0.876
3|4	60.93	37.69 – 98.50	<0.001	2.67	1.05 – 6.83	0.040
SyllStr-Open	55.28	36.30 – 86.16	<0.001	20.39	9.71 – 44.13	<0.001
SyllStr-sC	7.82	3.99 – 15.37	<0.001	3.37	0.58 – 28.44	0.205
LogFrequency	1.19	1.12 – 1.25	<0.001	1.23	1.05 – 1.46	0.016
Grapheme-E	1.14	0.84 – 1.55	0.412	0.84	0.36 – 2.02	0.694
Grapheme-O	0.20	0.14 – 0.28	<0.001	0.26	0.11 – 0.58	0.001
Grapheme-U	0.02	0.01 – 0.04	<0.001	1.15	0.34 – 4.76	0.830
WeightSi-Light	Irrelevant			2.51	1.31 – 4.90	0.006
AIC	2,167.078			369.182		

obstruents, but there does seem to be an interaction between the weight of the first syllable and vowel reduction in the second syllable. Finally, we could not find any effects of the nature of the coda (tested through Coda-Place and Coda-Manner). This was tested in the subset of words with a closed syllable and a vowel monograph (146 and 45 words), and neither variable was found to improve the models.

The variable Grapheme was tested on the subsets of 1,102 and 408 words with one of the four monographs ⟨a, e, o, u⟩ along with the variables used in the models reported above, with the exception of Foreign, which caused models to fail to converge. Grapheme was found to significantly improve models. The models are shown in Table 4, and the distribution of the data among open and closed syllables is shown in Figure 4 (vowels followed by /sC/ are left out for clarity, and because several categories have very small numbers of relevant forms).

What those results show is that, overall, ⟨o⟩ and ⟨u⟩ rarely represent reduced vowels, especially in initial pretonic position, while ⟨a⟩ and ⟨e⟩ often do. However, ⟨a⟩ patterns with ⟨o⟩ and ⟨u⟩ before ⟨rC⟩, where it is almost systematically realised as /ɑː/. It may only reduce in intertonic position, and only optionally. This observation, along with the greater resistance of ⟨o⟩ and ⟨u⟩ to vowel reduction, could be interpreted as a sign that back vowels resist more than front vowels. However, we should be cautious in interpreting the results regarding ⟨u⟩ because Wells (2008), our source of pronunciations for the intertonic data, tends to use the symbol /u/ more often than Jones (2006), our source for the initial pretonic data (as discussed in §3.1). The difference in behaviour for ⟨u⟩ that one may be tempted to see in Figure 3 may actually be an artefact of the different transcription systems used in the two dictionaries.

4.2. Prefixed words

This section deals with prefixed words, focusing on reduction in monosyllabic prefixes. Therefore, we are only concerned with initial pretonic position here ($n = 1,997$). This data set contains only vowel monographs; 1,271 words were coded as Opaque (e.g., *arise, confess, exploit*) while 726 were coded as Transparent (e.g., *co-author, desexualize, unaspirated*). We ran ordinal logistic regression with an additional variable, Transparency, to encode that difference. The results of that regression are shown in Table 5, and the distribution of the data is shown in Figure 5. Examples are shown in (7).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

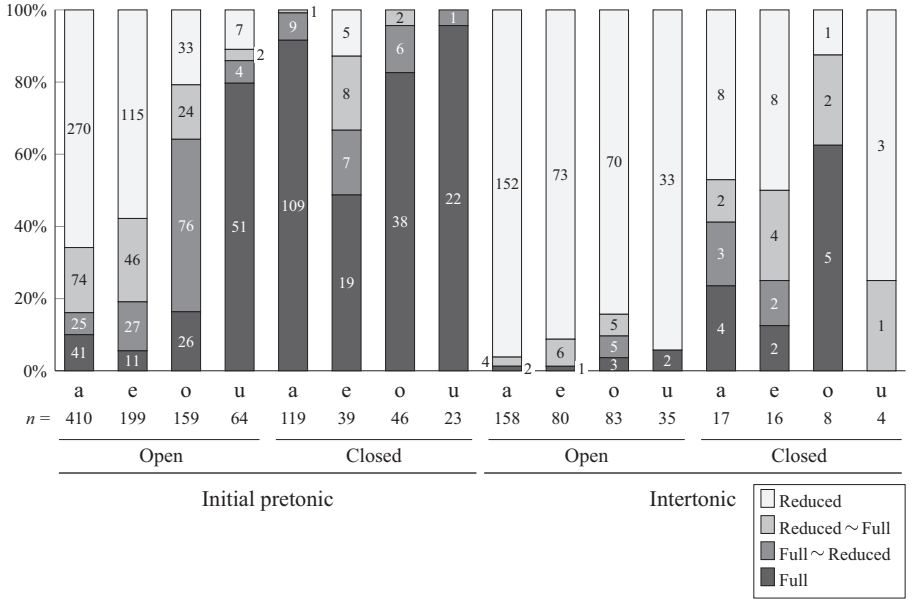


Figure 4. *Vowels found for depending on syllable structure and grapheme, among ⟨a, e, o, u⟩.*

Table 5. *Ordinal logistic regression for prefixed words.*

Predictor	Odds ratio	95% C.I.	*p*
1|2	0.31	0.22 – 0.44	<0.001
2|3	1.00	0.75 – 1.35	0.984
3|4	3.10	2.30 – 4.16	<0.001
SYLLSTR-OPEN	8.81	6.74 – 11.58	<0.001
SYLLSTR-SC	3.92	2.22 – 7.22	<0.001
LOGFREQUENCY	1.24	1.19 – 1.30	<0.001
TRANSPARENCY-TRANSPARENT	0.01	0.00 – 0.01	<0.001
AIC		2,134.460	

(7) a. *Opaque*

i. *Open:* *abide* /əˈbaɪd/
occult /əˈkʌlt/
dilate /daɪˈleɪt/ ~ /dɪˈleɪt/
diverse /daɪˈvɜːs/

ii. *Closed:* *complicit* /kəmˈplɪsɪt/
obstruct /əbˈstrʌkt/
extinct /ɪkˈstɪŋkt/ ~ /ɛkˈstɪŋkt/
ensign /ɛnˈsaɪn/

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

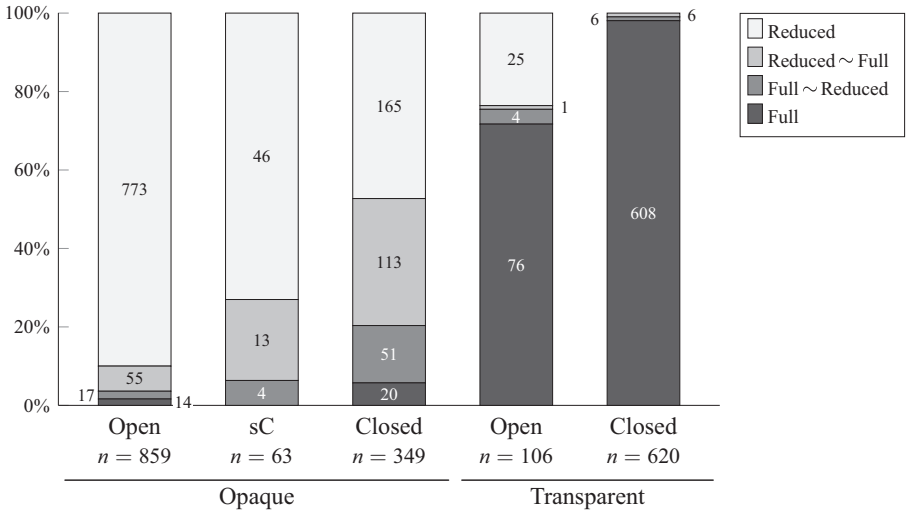


Figure 5. *Vowels found in the two types of prefixed words depending on syllable structure.*

b. *Transparent*

i.	*Open:*	*deform*	/dɪˈfɔːm/
		atonic	/eɪˈtɒnɪk/ ~ /aˈtɒnɪk/ ~ /əˈtɒnɪk/
		desexualize	/diːˈsɛkʃʊəlaɪz/
		unable	/ʌˈneɪbəl/
ii.	*Closed:*	*abnormal*	/abˈnɔːməl/ ~ /əbˈnɔːməl/
		subclassify	/sʌbˈklasɪfaɪ/
		transship	/transˈʃɪp/
		unlock	/ʌnˈlɒk/

These results confirm the effects of syllable structure and frequency in prefixed words, although a closer look at the data shows that transparent prefixed words do not appear to be sensitive to frequency, and almost systematically have full vowels, as shown in Figure 6. This is consistent with analyses which posit that in these words, the prefix is phonologically independent from the base (see fn. 3).

We observe a clear difference between opaque prefixed words and transparent prefixed words, but also that the reduction rates found here for opaque prefixed words differ strongly from those of non-prefixed words reported in the previous section. This confirms the numerous observations made in the literature regarding the reduction behaviour of opaque prefixes. Finally, it should be noted that this inventory displays considerably more stress variation than the rest of the data set: 56 opaque prefixed words have a variant in which the first vowel is stressed (25 for which the dictionary shows a secondary stress mark, 31 with a primary stress mark), while 426 transparent prefixed words (59%) have such a variant (421 with a possible secondary stress mark, 5 with a possible primary stress mark).

4.3. Stress-shifted suffixal derivatives

This section deals with suffixal derivatives in which the primary stress is shifted rightwards relative to its position in the corresponding base. The syllable of interest is stressed in the base, and so the bases have stress on the first syllable for derivatives where we are looking at the initial pretonic position (n = 590), or on the second syllable for intertonic position (n = 199).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

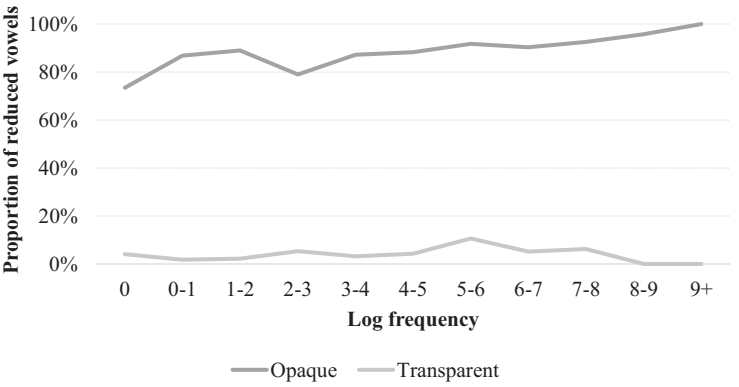


Figure 6. *Proportion of words with a reduced main pronunciation depending on the type of prefixed word.*

As in non-derived words, vowels represented by digraphs are mainly found in open syllables and are often full vowels. In derivatives, their numbers are too low to include them in statistical models: 23 in initial pretonic position and 5 in intertonic position. In intertonic position, there are only 11 items for which the vowel is followed by /sC/, and so we also excluded this configuration from statistical analyses. The following results are therefore based on 567 words for initial pretonic position and 184 words in intertonic position.

In order to test possible effects of base frequency, we tested four different configurations for the frequency variables alongside the other predictors: derivative frequency alone; absolute base and derivative frequency; relative frequency; and relative frequency with absolute derivative frequency. Then, we simplified the models following the procedure described in §3.3. As base and relative frequencies can be correlated, we systematically checked if this was a potential issue in our models using the Variance Inflation Factor (VIF; Zuur *et al.* 2010). This was implemented using the `vif` function of the `car` package, which identified no potentially harmful collinearity, as no variable ever had a VIF measure above 3.

The best models, shown in Table 6, reveal effects of SyllableStructure, LogFrequency and Morphology which are consistent with the results found for non-derived words. Those imply that a vowel is less likely to reduce if:

- the vowel is in a closed syllable;
- the word has a low frequency; or
- the vowel is not part of an opaque monosyllabic prefix.

However, in intertonic position, no effect of WeightSi can be identified in this data set. Now, turning to the variables that are specific to stress-shifted derivatives, we find effects of the characteristics of the vowel in the base and of variables meant to measure segmentability. For vowel characteristics, the results are not consistent across the two positions: we find an effect of VowelQuantity (short vowels being more likely to reduce than long vowels) and Back in initial pretonic position (vowels coded as [+back] are less likely to reduce than those coded [−back]). In intertonic position, we have an effect of High, as [+high] vowels reduce more than [−high] vowels. Examples illustrating the different configurations for VowelQuantity, Back and High are shown in (8) and (9).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Table 6. *Ordinal logistic regression for stress-shifting derivatives.*

	Initial pretonic			Intertonic		
Predictor	Odds ratio	95% C.I.	*p*	Odds ratio	95% C.I.	*p*
1|2	7.46	3.74 – 14.90	<0.001	0.44	0.17 – 1.15	0.093
2|3	29.91	14.31 – 62.55	<0.001	1.64	0.64 – 4.25	0.303
3|4	123.89	56.54 – 271.44	<0.001	4.63	1.74 – 12.31	0.002
SyllStr-Open	17.61	11.15 – 28.50	<0.001	7.66	4.00 – 15.11	<0.001
SyllStr-sC	0.30	0.08 – 0.89	0.047	Not enough data		
LogFrequency	1.17	1.09 – 1.26	<0.001	1.42	1.26 – 1.62	<0.001
LogFrequency-Base	1.09	1.01 – 1.18	0.026	No effect		
Morphology-Prefixed	16.46	8.91 – 31.31	<0.001	Irrelevant		
VowelQuantity-Short	2.16	1.44 – 3.26	<0.001	No effect		
Back-Yes	0.25	0.17 – 0.37	<0.001	No effect		
High-Yes	No effect			5.99	1.68 – 28.86	0.012
SemanticTransparency-Yes	0.67	0.46 – 0.96	0.031	0.25	0.10 – 0.61	0.003
WeightSi-Light	Irrelevant			No effect		
AIC	1,158.773			379.471		

(8) *Initial pretonic position*

a. *Long,* [+back]

/ɔː/ - /ɔː/	*balsam • balsamic* *author • authorial* *formal • formality*	/ɔː/ - /ə/	*ball • balloon*
/(j)uː/ - /(j)uː/	*brutal • brutality* *human • humanity* *Pluto • plutonian*	/(j)uː/ - /u/	*cupid • cupidity* *dual • duality* *fluid • fluidity*
/ɑː/ - /ɑː/	*harmony • harmonious* *plant • plantation* *sarcasm • sarcastic*	/ɑː/ - /ə/	*drama • dramatic*
/aɪ/ - /aɪ/	*cite • citation* *climate • climatic* *iamb • iambic*	/aɪ/ - /i/	*quiet • quiescence* *quiet • quiescent*

b. *Long,* [−back]

/eɪ/ - /eɪ/	*aphid • aphidian* *chaos • chaotic* *deist • deistic*	/eɪ/ - /ə/	*basic • basicity* *maniac • maniacal* *stable • stability*
/iː/ - /ɪ ~ iː/	*equal • equality* *feline • felinity* *theme • thematic*	/iː/ - /ɪ/	*demon • demoniac* *gene • genetic* *secret • secretive*

c. *Short,* [+back]

/ɒ/ - /ɒ/	*bombast • bombastic* *pompous • pomposity* *toxic • toxicity*	/ɒ/ - /ə/	*comedy • comedian* *concept • conceptual* *prodigy • prodigious*

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

/ʌ/ - /ʌ/	*custody* ⬩ *custodial* *judgment* ⬩ *judgmental* *sulphur* ⬩ *sulphuric*	/ʌ/ - /ə/	*courage* ⬩ *courageous* *substance* ⬩ *substantial* *substance* ⬩ *substantiate*

d. *Short,* [−back]

/a/ - /a/	*active* ⬩ *activity* *ancestor* ⬩ *ancestral* *fragment* ⬩ *fragmental*	/a/ - /ə/	*acid* ⬩ *acidic* *galaxy* ⬩ *galactic* *malice* ⬩ *malicious*
/ɛ/ - /ɛ/	*central* ⬩ *centrality* *ethnic* ⬩ *ethnicity* *Neptune* ⬩ *neptunian*	/ɛ/ - /ɪ ~ ə/	*definite* ⬩ *definitive* *edit* ⬩ *edition* *remedy* ⬩ *remedial*

(9) *Intertonic position*

a. [+high]

/uː/ - /u/	*acoustic* ⬩ *acoustician* *compute* ⬩ *computation* *refute* ⬩ *refutation*
/iː/ - /ə ~ ɪ/	*compete* ⬩ *competition* *Mycenae* ⬩ *Mycenaean*

b. [−high]

/ɔː/ - /ɔː/	*exalt* ⬩ *exaltation* *exhort* ⬩ *exhortation* *debauch* ⬩ *debauchee*	/ɔː/ - /ə/	*inform* ⬩ *information* *notorious* ⬩ *notoriety* *restore* ⬩ *restoration*
/a/ - /a/	*contractile* ⬩ *contractility* *recant* ⬩ *recantation* *refractive* ⬩ *refractivity*	/a/ - /ə/	*academy* ⬩ *academic* *Japan* ⬩ *Japanese* *magnanimous* ⬩ *magnanimity*
/ɛ/ - /ɛ/	*accept* ⬩ *acceptation* *director* ⬩ *directorial* *molest* ⬩ *molestation*	/ɛ/ - /ə ~ ɪ/	*gazette* ⬩ *gazetteer* *ingenuous* ⬩ *ingenuity* *Vienna* ⬩ *Viennese*

As can be seen in (9), the feature [±high] here distinguishes /iː/ and /(j)uː/ from other vowels (remember that /ɪ/ and /ʊ/ are almost systematically excluded from the data), and most [+high] cases are words in which the base has /(j)uː/. We saw in §3.1 that there are disagreements between our two sources regarding that vowel, and so those results should probably be interpreted with caution. In order to see if this variable affected our final model, we tried excluding HIGH from the start of the model simplification procedure, and then went on with gradually removing predictors that do not satisfy the conditions described in §3.3. No other predictors made it to the final model except those that are present alongside HIGH in Table 6, and the model has a higher AIC than that in Table 6 (385.4849 *vs.* 379.471). Therefore, HIGH does not affect which predictors are included in the model. The results regarding VOWELQUANTITY and BACK in initial pretonic position seem more robust, as they contrast more different vowels. The effect of the feature [±back] is somewhat consistent with the results reported for GRAPHEME in non-derived words, in which we found that ⟨o⟩ and ⟨u⟩ (which always represent [+back] vowels) are less often reduced than ⟨a⟩ and ⟨e⟩ (which rarely represent [+back] vowels).

The results are also unclear regarding the variables meant to measure segmentability. We do find an effect of SEMANTICTRANSPARENCY in both positions, but the effects of base frequency are quite intriguing. In initial pretonic position, we do find an effect of base frequency, and the best model (reported in Table 6) is one which includes absolute base and derivative frequencies. However, the effect is in the opposite direction from what the segmentability hypothesis predicts: the more frequent the base is, the more likely the vowel is to be reduced. In that same position, we also found a model

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Table 7. *Ordinal logistic regression for all words with one of the four monographs ⟨a, e, o, u⟩.*

	Initial pretonic			Intertonic		
Predictor	Odds ratio	95% C.I.	*p*	Odds ratio	95% C.I.	*p*
1|2	2.93	2.28 – 3.77	<0.001	0.27	0.14 – 0.52	<0.001
2|3	9.57	7.36 – 12.45	<0.001	0.74	0.39 – 1.41	0.363
3|4	28.97	21.99 – 38.17	<0.001	2.01	1.05 – 3.84	0.034
SyllableStructure-Open	16.25	13.16 – 20.17	<0.001	16.79	9.84 –29.23	<0.001
SyllableStructure-sC	2.74	1.87 – 4.02	<0.001	0.62	0.23 – 1.66	0.344
LogFrequency	1.23	1.19 – 1.27	<0.001	1.35	1.22 – 1.50	<0.001
Grapheme-E	0.80	0.66 – 0.98	0.035	0.97	0.53 – 1.79	0.929
Grapheme-O	0.40	0.32 – 0.49	<0.001	0.33	0.18 – 0.59	<0.001
Grapheme-U	0.09	0.06 – 0.13	<0.001	1.70	0.77 – 3.95	0.201
Derived-Yes	0.58	0.47 – 0.70	<0.001	0.16	0.10 – 0.27	<0.001
Morphology-Prefixed	13.72	11.03 – 17.17	<0.001	Irrelevant		
AIC	5,093.552			743.374		

in which relative frequency is a good predictor of vowel reduction, but it requires the exclusion of SemanticTransparency, and the overall AIC of the model is far higher than the one reported in Table 6 (1,195.019 *vs.* 1,158.773). In this model, the direction of the effect for relative frequency goes in the expected direction: the more frequent the derivative relative to the base (so the higher relative frequency), the more likely the vowel is to be reduced. In intertonic position, no base frequency variables (absolute or relative) made it into the final model, and so base frequency cannot be considered a good predictor of vowel reduction in that position in stress-shifted derivatives.

4.4. Non-derived words vs. stress-shifted derivatives

The results reported in the previous sections show no clear effects of segmentability. However, we may try to compare non-derived words to stress-shifted derivatives in a more categorical way so as to establish whether or not the existence of a base with primary stress on the relevant vowel reduces the probability of reduction (as in the *condensation–compensation* contrast), as claimed in the literature reviewed in §2.3. The comparison between the two data sets is complicated because the variables encoding vowel properties are not the same. Indeed, in non-derived words, as we do not always have access to a surface vowel, we used spelling as a way to approach how different types of vowels behave. Therefore, we ran regression analyses using the Grapheme variable, as it is the only information that we can compare across the different data sets. We ran these analyses using SyllableStructure, LogFrequency, Morphology (for initial pretonic position only) and an additional variable Derived with two possible values (Yes or No). Regarding the effects of vowel characteristics, we will only report the results for the subsets of words that contain one of the four monographs ⟨a, e, o, u⟩, but we also ran models with the full data set and found similar results and a significant effect of Spelling. Those data sets contain 3,617 words in initial position and 587 in intertonic position. The results of the regression analysis are shown in Table 7.

The results shown in Table 7 confirm once again the effects of SyllableStructure, LogFrequency, Grapheme and Morphology. Moreover, they show a strong effect of Derived in both positions, which shows that having a base with stress on the relevant vowel significantly reduces the chances of reduction in the related derivative. The distribution of the data is shown in Figure 7 for non-prefixed words and in Figure 8 for prefixed words.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

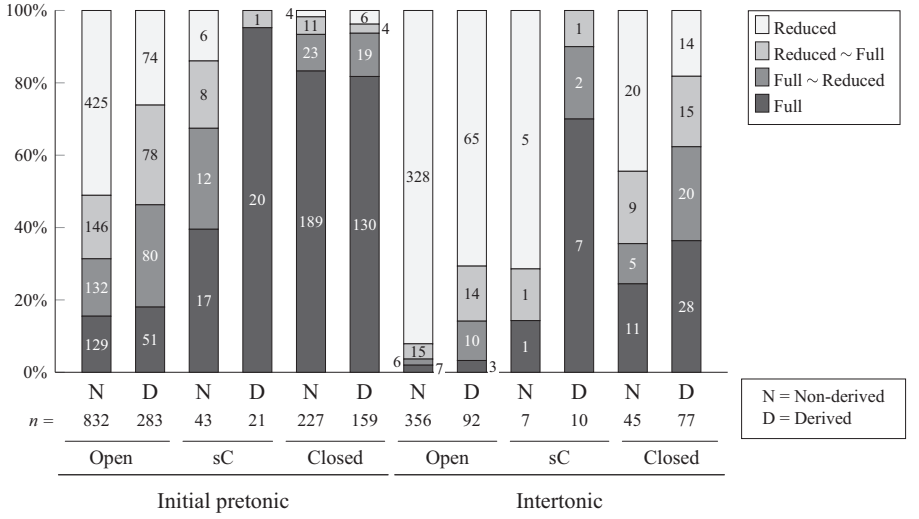


Figure 7. *Vowels found for monographs ⟨a, e, o, u⟩ in non-prefixed derived and non-derived words, for both positions and depending on syllable structure.*

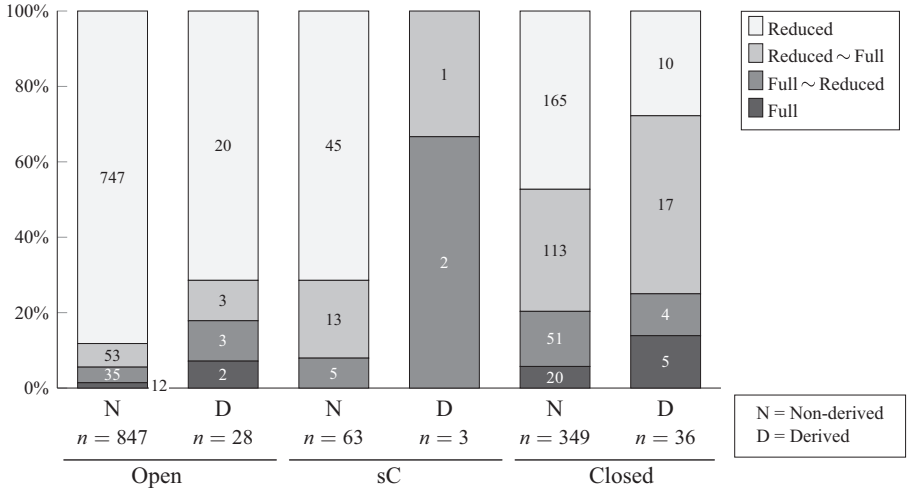


Figure 8. *Vowels found for monographs ⟨a, e, o, u⟩ in derived and non-derived words containing an opaque prefix, for both positions and depending on syllable structure.*

Two main things can be observed in these figures. First, although we did not find any clear segmentability effects in derivatives, we do find a significant difference between derived and non-derived words in all configurations. Second, the behaviour of vowels preceding /sC/ clusters patterns with that of vowels in closed syllables in derived words, with very little reduction, whereas the reduction rates found in non-derived words before /sC/ were between those for open syllables and those for closed syllables. Let us now turn to the interpretation of these results.

5. Discussion

In this section, we summarise the evidence that our study has provided and how it relates to the literature reviewed in §2, and then we turn to the implications of our results.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

5.1. The determining factors of vowel reduction

The study we have just presented is the most comprehensive to date on the issue of vowel reduction in English, as it is the only one to have tested such a variety of factors on so large a data set. We found that certain factors were indeed significant predictors of vowel reduction, while no evidence could be found for others, and we found evidence for new factors not previously proposed in the literature.

First, our results confirm the importance of position, as we observe considerably more vowel reduction in intertonic position than in initial pretonic position. We also report clear evidence of the role played by syllable structure, with vowels in closed syllables being less often reduced than vowels in open syllables, and those followed by /sC/ clusters showing a somewhat intermediate behaviour in non-derived words. We have strong evidence of the role played by frequency, and we found that it can be confirmed independently of effects attributable to foreignness. We also have clear evidence that opaque prefixes favour reduction in initial pretonic position, and that words containing such prefixes are different in that regard from both non-prefixed words and transparent prefixed words. This latter point will be discussed in §5.4.

Then, there is weak evidence for other factors. Vowels spelled with digraphs appear to reduce less than those spelled with monographs, although we cannot test whether this is attributable to the fact that digraphs may represent certain types of vowels that happen to reduce less. The only data set in which it might have been possible to do this is in stress-shifted derivatives, as both phonological and orthographic source vowels are available, but we did not have enough data for digraphs to include them in the analysis. We have some evidence for an influence of the weight of the first syllable on vowel reduction in intertonic position, although this factor was found to be a significant predictor only in non-derived words. We have evidence from the derived-words data set that long vowels reduce less than short ones, in line with Chomsky & Halle's (1968) claim. We have evidence regarding the role played by the existence of a base in which the vowel is stressed, as stress-shifted derivatives reduce significantly less than non-derived words, but no clear segmentability effects could be found. We discuss this latter point in §5.2.

However, we found no effects of the nature of the coda in closed syllables. The effects of segmentability are contradictory, as we do find an effect of semantic transparency in both positions, but we only find an effect of base frequency in initial pretonic position, and it goes in the opposite direction from that predicted by the segmentability hypothesis.

Finally, our study has allowed us to bring forward three observations which have not been made previously in the literature. First, we confirmed Dabouis & Fournier's (2022) hypothesis that words which can be identified as foreign undergo less reduction than the rest of the lexicon, even though this is a small difference. Second, our results show that, in initial pretonic position, derived words for which the vowel in the base is [+back] undergo less reduction than those for which the vowel in the base is [−back], and this is consistent with the observations made in non-derived words, where it was found that orthographic ⟨o⟩ and ⟨u⟩, which are often realised as [+back] vowels, reduce less than ⟨a⟩ and ⟨e⟩, which are almost never realised as [+back]. Those results are consistent with those reported by Tokar (2019), who finds that ⟨o⟩ reduces less than ⟨a⟩ in open syllables in initial pretonic position. We also found an effect of [±high] in intertonic position, but, as discussed above, that result may not be reliable. Finally, we found a difference between derived and non-derived words regarding vowels followed by /sC/, for which we propose an interpretation in §5.3.

5.2. Preservation and segmentability

One of the main aims of this article was to establish whether there is empirical support for the kind of vowel preservation described by Chomsky & Halle (1968) in their discussion of the *condensation–compensation* pair. Our results confirm that there is a difference between words that have a base in which the vowel bears a stress and those that do not, as vowels are less often reduced in the former than in the latter. As pointed out by many before us, this is not a categorical difference. This is not really

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

surprising, considering that our results show that vowel reduction is a highly variable phenomenon that is determined by many different factors, and so the requirement that the vowel of the base should be preserved appears to be one constraint among others.

Another finding regarding identity relationships between words is the absence of clear evidence of segmentability effects, which we tested through (absolute and relative) base and derivative frequencies and through semantic transparency. We do find effects of semantic transparency, but we find effects of base frequency only in initial pretonic position, and the effect of the base is opposite from the prediction of the segmentability hypothesis. The hypothesis would predict that a higher base frequency should reduce the likelihood of vowel reduction. Thus, our results are different from those reported by Kraska-Szlenk (2007) (who uses only a small sample of words in intertonic position, in which the vowel may only be followed by a sonorant), but are similar to those reported by Hammond. Hammond (2003: 44) reports the same effect of base frequency in intertonic position, and proposes that 'the frequency of a complex derived form is a partial function of the frequency of its part'. This means that we would expect more reduction if the cumulated frequency of the base and its derivative is higher. We tested that idea by fitting models using cumulated base and derivative frequencies (as *ln*(derivative frequency + 1) + *ln*(base frequency + 1)), but those models are not better than the ones reported in §4.3. It is possible that, as Hammond suggests, we actually need to take the overall frequency of the components into account, which would require us to consider the frequencies of the whole morphological family, not just those of the base and the derivative (see the last paragraph of this section). It is also possible that fusing the two frequencies together is not a good option, as their effects may differ in magnitude, since the local base often has a closer formal and semantic connection to the derivative.

One possibility is that the dictionary data are not fine-grained enough to allow for the clear identification of segmentability effects. Arndt-Lappe & Dabouis (in preparation) report a study on weak stress preservation with both dictionary data and speech data; they do not find any segmentability effects in the former, but they do find such effects in the latter. Therefore, another possibility is that segmentability effects are difficult to detect in dictionary data for this process, but might be more easily detected in speech data. Arndt-Lappe & Dabouis's speech data also includes derivatives with a higher frequency than those of their dictionary data, which may also explain the difference: frequency effects might be visible only in certain frequency ranges. Therefore, future research could test the same variables as those used in this study using speech data to establish whether or not any effects of segmentability can be observed.

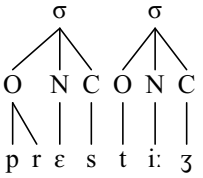
Finally, we could consider base–derivative identity from another perspective than that of segmentability. The hypothesis relies on dual-route race models of lexical access, in which only embedded bases are considered. Most approaches use the local base (e.g., *connective* • *connectivity*), but some consider more deeply embedded bases (e.g., *connect* • *connective* • *connectivity*; Bermúdez-Otero 2007; Dabouis 2019). Other approaches, notably Lexical Conservatism (Steriade 1997; Steriade & Stanton 2020; Breiss 2021), assume that other words from the morphological family may be used as bases. Analogy-based frameworks (e.g., Arndt-Lappe 2015) also do not impose restrictions on containment or locality of potential analogues (i.e., words used as models in the computation of another word). Therefore, it is possible that in order to account for the difference between stress-shifted derivatives and non-derived words, we need to include other words from the morphological family of the derivative. Future studies will have to establish whether that can be useful, and, if other bases are relevant, whether only one should be considered or whether there can be simultaneous influences from multiple bases. For example, can vowel reduction be better predicted if, instead of using the frequency of the local base, we use the cumulated frequencies of all the words in the morphological family of the derivative that have stress on the relevant vowel (e.g., *condúct, condúctive, condúctance, condúction, condúctor, condúctress* for *còndu̲ctívity*)? Exploratory work conducted by Dabouis (2023b) on vowel reduction in English suggests that this can be a fruitful area to explore.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

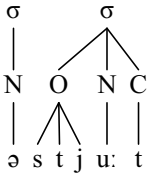
5.3. /sC/ clusters

Our results show that in non-derived words, the rates of vowel reduction for vowels followed by /sC/ clusters are between those observed for vowels in open syllables and those observed for vowels in closed syllables, but we found them to pattern with closed syllables in derived words. The syllabification of these clusters is notoriously controversial, as they can appear in word-initial position and thus may be taken to be a well-formed onset, although the fact that they usually have falling sonority suggests that /s/ should be analysed as a coda. One possibility is that these clusters sometimes syllabify heterosyllabically and sometimes tautosyllabically, and we could take our results to mean that words with full vowels are those in which /s/ is syllabified as a coda (as in (10a)), while those with reduced vowels are those in which /s/ is syllabified as part of a complex onset (as in (10b)).[12]

(10) a. *Heterosyllabic parsing:* prestige



b. *Tautosyllabic parsing:* astute



Considering that in derived words, vowels followed by /sC/ pattern with vowels in closed syllables, we can assume that these words have structures such as that in (10a). Our hypothesis as to why this would be the case is that syllabification is inherited from the base, and this is possible if we accept two assumptions: that stressed vowels force a following /sC/ cluster to be syllabified heterosyllabically, and that prosodic structure (here, syllable structure) may be partially preserved, even if there are modifications in prosodic structure elsewhere in the word when an stress-affecting suffix is attached. The former can be taken to be a form of coda maximization (Wells 1990) or a requirement that stressed syllables be heavy (Duanmu 2015), while the latter is similar to Davis's (2005) analysis of the pair *capitalistic vs. militaristic.* Davis assumes that /t/-flapping is phonologically unexpected in *capitalistic* and is attributable to preservation (analysed as 'paradigm uniformity') of the prosodic structure of *capital.* To summarise our analysis with an example, we assume that the /sC/ cluster in *plastic* is syllabified heterosyllabically, and that /s/ is maintained in the coda of *plasticity,* thus explaining why the first vowel remains non-reduced /a/. This is to be contrasted with non-derived words such as those in (10a), for which the syllabification of /sC/ may vary.

5.4. Opaque morphology and phonology

The observation that semantically opaque prefixes behave differently from both transparent prefixed words and non-prefixed words needs to be commented upon, as some might be reluctant to include units such as *ad-, con-, -mit* or *-ceive* among possibly relevant morphological units, since they do not fit classical definitions of the morpheme. Let us directly compare initial pretonic vowels in non-prefixed words (seen in §4.1) with those in words containing an opaque prefix (§4.2). If we consider only the main pronunciation and only open and closed syllable for a clearer comparison, the difference is striking, as can be seen in Figure 9.

Indeed, we can see that words with an opaque prefix have a reduced vowel in 97% of words with an open syllable, as opposed to 69% for non-prefixed words. This difference is even more striking in closed syllables, with 80% of vowels reduced in words with opaque prefixes, as opposed to 7% in non-prefixed words.

This is not the only phonological behaviour that distinguishes opaquely prefixed words from morphologically simple words: there are also differences in stress placement in verbs (Chomsky &

[12] An alternative suggested to us by one of the editors is that the intermediate behaviour of /sC/ clusters might be taken as evidence that the /s/ is ambisyllabic (Kahn 1976).

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

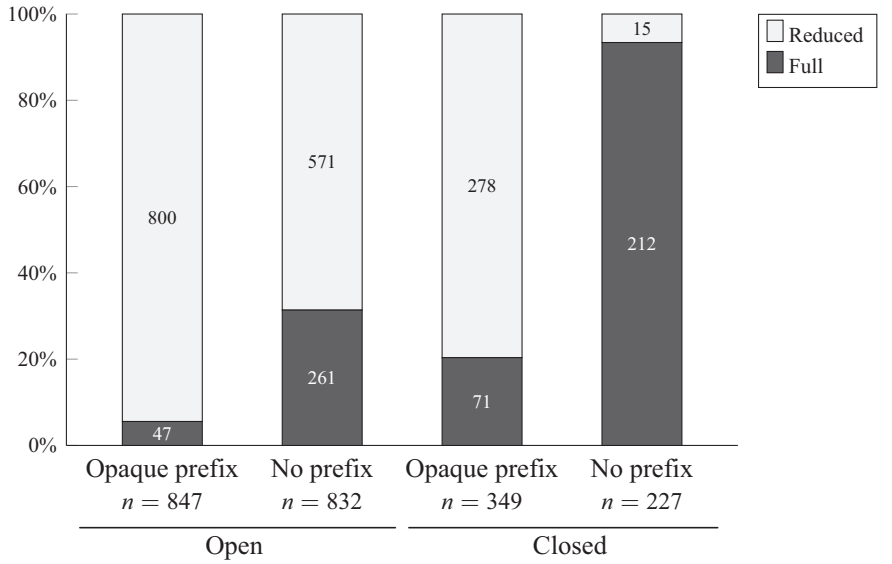


Figure 9. *Main pronunciation of the vowel in the initial pretonic syllable of non-derived words which contain an opaque monosyllabic prefix and those which do not.*

Halle 1968; Guierre 1979; Dabouis & Fournier 2023), the diachronic evolution of stress in verb–noun pairs (Sonderegger & Niyogi 2013), stress preservation (Dabouis 2019) and the phonotactics of word-medial consonant clusters (Guierre 1990; Hammond 1999). There is also considerable evidence from psycholinguistics that such words are analysed and stored as complex words, from lexical decision tasks (Taft & Forster 1975; Taft *et al.* 1986; Taft 1994; Forster & Azuma 2000; Pastizzo & Feldman 2004), reading studies (Rastle & Coltheart 2000; Ktori *et al.* 2016, 2018) and ERP (McKinnon *et al.* 2003). The structural property of these words that has most often been put forward to account for the emergence of units such as *-mit* or *ad-* in the absence of clear semantic support is the distributional recurrence of most of the prefixes and roots involved (Taft 1994; Fournier 1996; Forster & Azuma 2000), but there are also other possible elements contributing to this emergence (see Dabouis & Fournier 2025 for a full discussion). Our study therefore provides additional evidence that such units are relevant for phonology, and that phonological theories should be able to refer to opaque morphological structure.

Raffelsiefen (2007) argues that the difference in initial pretonic vowel reduction between prefixed and non-prefixed words has to do with their syntactic category: nouns do not have reduced vowels, but verbs do. However, she only uses disyllabic examples to support her claim, and Dabouis & Fournier (2025) show that it is impossible to tease that claim apart from the alternative that the difference in reduction behaviour is attributable to prefixation in disyllables. Our data show that the difference between prefixed and non-prefixed words is highly significant, even though the latter are often nouns or adjectives and can be longer than two syllables. The difference persists in stress-shifted derivatives, with the proportion of full vowels being lower in prefixed words in both open syllables (48% *vs.* 80%) and closed ones (9% *vs.* 73%). This last observation is potentially a problem for cyclic theories such as Stratal Phonology that assume a form of bracket erasure, such that morphosyntactic information from earlier cycles is not visible to the phonology. Such a model would not predict, for example, that the vowel of *conceptual* (⋆ *concept*) should reduce while that of *pomposity* (⋆ *pompous*) should not, because the information that *conceptual* is prefixed should not be available: this information should be lost after the phonological representation of *concept* has been computed (and lexically stored).

One aspect of the phonology of those opaque prefixes which remains to be explored is the possible effects of the frequency of each prefix. As mentioned in §2.4, Hammond (2003) suggests that the high

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

frequency of those prefixes explains why they are so often reduced. Thus, future research could extend the present study by integrating a measure of prefix frequency, and Hammond's claim would predict that prefixes with higher frequency reduce more than those with lower frequency.

6. Conclusion

In this article, we have defined our object of study, vowel reduction and preservation in English, and reviewed the proposals made in the literature regarding what influences these processes. We have presented the most comprehensive study to date on the issue using dictionary data, with control data sets to establish how reduction functions in the absence of preservation effects. Our results show which structural, morphological and lexical factors influence vowel reduction and confirm that the existence of a base in which the vowel is stressed disfavours vowel reduction, although we have inconsistent results regarding segmentability. Our results have also shown that vowels followed by /sC/ clusters pattern with those in closed syllables in stress-shifted derivatives, but not in non-derived words, and we have suggested that this could be attributed to preservation of syllabic structure from the base, where we assume that these clusters are parsed heterosyllabically. Finally, we saw that our results fall in line with a number of previous studies dealing with semantically opaque prefixes and support the idea that these units are relevant for phonological computation.

Data availability statement. The full data sets and R scripts used in this article are available on OSF: https://osf.io/qbcnv/.

Acknowledgements. We would like to thank the audiences of the 2018 and 2019 Manchester Phonology Meeting, the 2019 PAC conference and the 2018 meeting of the French Phonology Network (RFP), in which earlier versions of this work were presented. We would also like to thank our collaborators on the first steps of this project, Nicola Lampitelli and Guillaume Enguehard, as well as Ricardo Bermúdez-Otero for his stimulating comments and suggestions. Finally, we thank two anonymous reviewers as well as the editors of this volume for their constructive criticism and suggestions which have allowed us to improve the article significantly. All errors are ours alone.

Funding statement. This research was supported by funding from the ANR (ANR-21-FRAL-0001-01) and the DFG (AR 676/3-1) for the ERSaF project (English Root Stress across Frameworks).

Competing interests. The authors declare no competing interests.

References

Arndt-Lappe, Sabine (2015). Word-formation and analogy. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.) *Word-formation: an international handbook of the languages of Europe*, volume **2**. Berlin: de Gruyter Mouton, 822–841.

Arndt-Lappe, Sabine & Quentin Dabouis (in preparation). *Secondary stress, lexical storage, and morphological structure: new evidence from dictionary and speech data*. Ms, Universität Trier and Université Clermont Auvergne.

Baayen, R. Harald (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Backley, Phillip (2011). *An introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Ballier, Nicolas, Jean-Michel Fournier, Anne Przewozny & Eiji Yamada (eds.) (2023). *New perspectives on English word stress*. Edinburgh: Edinburgh University Press.

Ballier, Nicolas & Philippe Martin (2010). Corrélats prosodiques et acoustiques de la syllabification: le cas du français et de l'anglais. *Paper presented at the 8th Meeting of the French Phonology Network (Réseau Français de Phonologie)*, University of Orléans, July 2010.

Bauer, Laurie, Rochelle Lieber & Ingo Plag (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* **60**, 92–111.

Bermúdez-Otero, Ricardo (2007). *On the nature of the cycle*. Paper presented at the 15th Manchester Phonology Meeting, University of Manchester, May 2007. Handout available online at http://www.bermudez-otero.com/15mfm.pdf.

Bermúdez-Otero, Ricardo (2012). The architecture of grammar and the division of labor in exponence. In Jochen Trommer (ed.) *The morphology and phonology of exponence*. Oxford: Oxford University Press, 8–83.

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

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Booij, Geert & Jerzy Rubach (1984). Morphological and prosodic domains in lexical phonology. *Phonology Yearbook* **1**, 1–27.
Breiss, Canaan (2021). *Lexical Conservatism in phonology: theory, experiments, and computational modeling*. PhD dissertation, University of California, Los Angeles.
Burzio, Luigi (1994). *Principles of English stress*. Cambridge: Cambridge University Press.
Burzio, Luigi (2007). Phonology and phonetics of English stress and vowel reduction. *Language Sciences* **29**, 154–176.
Carney, Edward (1994). *A survey of English spelling*. London: Routledge.
Carr, Philip, Jacques Durand & Colin J. Ewen (eds.) (2005). *Headhood, elements, specification and contrastivity: phonological papers in honour of John Anderson*. Amsterdam: Benjamins.
Cho, Hye-Sun (2004). Frequency and stress preservation: encoding *Frequency* in OT. *SNU Working Papers in English Language and Linguistics* **3**, 187–203.
Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.
Clopper, Cynthia G. & Rory Turnbull (2018). Exploring variation in phonetic reduction: linguistic, social, and cognitive factors. In Francesco Cangemi, Meghan Clayards, Oliver Nieburh, Barbara Schuppler & Margaret Zelllers (eds.) *Rethinking reduction: interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Berlin: de Gruyter Mouton, 25–72.
Collie, Sarah (2007). *English stress preservation and Stratal Optimality Theory*. PhD dissertation, University of Edinburgh.
Collie, Sarah (2008). English stress preservation: the case for 'fake cyclicity'. *English Language and Linguistics* **12**, 505–532.
Cruttenden, Alan (2014). *Gimson's pronunciation of English*. 8th edition. London: Routledge.
Dabouis, Quentin (2016). *L'accent secondaire en anglais britannique contemporain*. PhD dissertation, University of Tours.
Dabouis, Quentin (2019). When accent preservation leads to clash. *English Language and Linguistics* **23**, 363–404.
Dabouis, Quentin (2020). Secondary stress in contemporary British English: an overview. *Anglophonia* **30**.
Dabouis, Quentin (2023a). English phonology and the literate speaker: some implications for lexical stress. In *Ballier et al.* (**2023**), 117–153.
Dabouis, Quentin (2023b). Morphological gradience in phonology. *Paper presented at the 30th Manchester Phonology Meeting*, University of Manchester, May 2023.
Dabouis, Quentin, Guillaume Enguehard, Jean-Michel Fournier & Nicola Lampitelli (2020). The English "Arab Rule" without feet. *Acta Linguistica Academica* **1**, 121–134.
Dabouis, Quentin & Jean-Michel Fournier (2023). The stress patterns of English verbs: syllable weight or morphology? In Ballier *et al*. (2023), 154–191.
Dabouis, Quentin & Jean-Michel Fournier (2025). Opaque morphology and phonology: historical prefixes in English. *Journal of Linguistics* **61**, 231–263.
Dabouis, Quentin & Pierre Fournier (2022). English phonologieS? In Viviane Arigne & Christiane Rocq-Migette (eds.) *Modèles et modélisation en linguistique / models and modelisation in linguistics*. Brussels: Peter Lang, 215–258.
Dabouis, Quentin, Olivier Glain & Sylvain Navarro (2023). Patterns of /r/ gemination in British and American English: a comparative study. Paper presented at the 16th PAC conference, Université Paris Nanterre, April 2023. Slides available online at https://hal.science/hal-04070317.
Dahak, Anissa (2006). Quel statut pour un corpus 'oral' écrit? Le dictionnaire de prononciation comme corpus (quasi-)exhaustif de la langue. In *Actes des IXèmes Rencontres Jeunes Chercheurs de l'École Doctorale 268 'Langage et Langues'*. Paris: Université de la Sorbonne Nouvelle, 1–5.
Dahak, Anissa (2009). Vowels in inter-tonic syllables. In Carlos Prado-Alonso, Lidia Gómez-García, Iria Pastor-Gómez & David Tizón-Couto (eds.) *New trends and methodologies in applied English language research: diachronic, diatopic and contrastive studies*. Bern: Peter Lang, 131–151.
Dahak, Anissa (2011). *Étude diachronique, phonologique et morphologique des syllabes inaccentuées en anglais contemporain*. PhD dissertation, Université de Paris Diderot.
Davis, Stuart (2005). Capitalistic v. militaristic: the paradigm uniformity effect reconsidered. In Laura J. Downing, T. Alan Hall & Renate Raffelsiefen (eds.) *Paradigms in phonological theory*. Oxford: Oxford University Press, 107–121.
Deschamps, Alain (1994). *De l'écrit à l'oral et de l'oral à l'écrit*. Paris: Ophrys.
Deschamps, Alain, Jean-Louis Duchet, Jean-Michel Fournier & Michael O'Neil (2004). *English phonology and graphophonemics*. Paris: Ophrys.
Duanmu, San (2015). Onset and the weight-stress principle in English. In Bo Hong, Fuxiang Wu & Chaofen Sun (eds.) *Linguistic essays in honor of Professor Tsu-Lin Mei on his 80th birthday*. Beijing: Capital Normal University Press, 295–338.
Durand, Jacques (2005). Tense/lax, the vowel system of English and phonological theory. In Carr et al. (2005), 77–90.
Durand, Jacques & Eiji Yamada (2023). On the treatment of English word stress within the generative tradition: history, concepts and debates. In Ballier *et al.* (2023), 6–52.
Engemann, Marie & Ingo Plag (2021). Phonetic reduction and paradigm uniformity effects in spontaneous speech. *The Mental Lexicon* **16**, 165–198.
Fidelholtz, James L. (1966). *Vowel reduction in English*. Ms, University of Maryland.
Fidelholtz, James L. (1975). Word frequency and vowel reduction in English. *CLS* **11**, 200–213.
Forster, Kenneth I. & Tamiko Azuma (2000). Masked priming for prefixed words with bound stems: does submit prime permit? *Language and Cognitive Processes* **15**, 539–561.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Fournier, Jean-Michel (1996). La reconnaissance morphologique. In A. Deschamps & J.-L. Duchet (eds.) *8ème Colloque d'Avril sur l'anglais oral*. Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV, 45–75.
Fournier, Jean-Michel (2010). *Manuel d'anglais oral*. Paris: Ophrys.
Fudge, Erik (1984). *English word stress*. London: Allen & Unwin.
Giegerich, Heinz J. (1999). *Lexical strata in English: morphological causes, phonological effects*. Cambridge: Cambridge University Press.
Goad, Heather (2012). sC clusters are (almost always) coda-initial. *The Linguistic Review* **29**, 335–373.
Greenwell, Brandon M., Andrew J. McCarthy, Bradley C. Boehmke & Dungang Liu (2018). Residuals and diagnostics for binary and ordinal regression models: an introduction to the sure package. *R Journal* **10**, 381–394.
Guierre, Lionel (1979). *Essai sur l'accentuation en anglais contemporain: éléments pour une synthèse*. PhD dissertation, Université Paris VII.
Guierre, Lionel (1990). Mots composés anglais et agrégats consonantiques. In J.-L. Duchet, J.-M. Fournier, J. Humbley & P. Larreya (eds.) *5ème Colloque d'Avril sur l'anglais oral*. Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV, 59–72.
Halle, Morris (1973). Stress rules in English: a new version. *LI* **4**, 451–464.
Halle, Morris & Samuel Keyser (1971). *English stress: its form, its growth, and its role in verse*. New York: Harper & Row.
Halle, Morris & Jean-Roger Vergnaud (1987). *An essay on stress*. Cambridge, MA: MIT Press.
Hammond, Michael (1999). *The phonology of English: a Prosodic Optimality-Theoretic approach*. Oxford: Oxford University Press.
Hammond, Michael (2003). *Frequency, cyclicity, and optimality*. Paper presented at the Second International Korean Phonology Conference, Seoul National University, June 2003. Slides archived at https://web.archive.org/web/20060903213233/http://www.u.arizona.edu/~hammond/kslides.pdf.
Harris, John (1994). *English sound structure*. Oxford: Blackwell.
Harris, John (2005). Reduction as information loss. In Carr *et al.* (2005), 119–132.
Harris, John & Edmund Gussmann (1998). Final codas: why the west was wrong. In Eugeniusz Cyran (ed.) *Structure and interpretation: studies in phonology*. Lublin: Folium, 139–162.
Hay, Jennifer (2001). Lexical frequency in morphology: is everything relative? *Linguistics* **39**, 1041–1070.
Hay, Jennifer (2003). *Causes and consequences of word structures*. London: Routledge.
Hay, Jennifer & R. Harald Baayen (2002). Parsing and productivity. In Geert Booij & Jaap Marle (eds.) *Yearbook of morphology 2001*. Dordrecht: Kluwer, 203–235.
Hayes, Bruce (1982). Extrametricality and English stress. *LI* **13**, 227–276.
van Heuven, Walter J.B., Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* **67**, 1176–1190.
Jackendoff, Ray (1975). Morphological and semantic regularities in the lexicon. *Lg* **51**, 639–671.
Jensen, John T. (2022). *The lexical and metrical phonology of English: the legacy of* The sound pattern of English. Cambridge: Cambridge University Press.
Jones, Daniel (2006). *Cambridge English pronouncing dictionary*. 17th edition. Cambridge: Cambridge University Press.
Kahn, Daniel (1976). *Syllable-based generalizations in English phonology*. PhD dissertation, Massachusetts Institute of Technology.
Kraska-Szlenk, Iwona (2007). *Analogy: the relation between lexicon and grammar*. Munich: LINCOM Europa.
Ktori, Maria, Petroula Mousikou & Kathleen Rastle (2018). Cues to stress assignment in reading aloud. *Journal of Experimental Psychology* **147**, 36–61.
Ktori, Maria, Jeremy J. Tree, Petroula Mousikou, Max Coltheart & Kathleen Rastle (2016). Prefixes repel stress in reading aloud: evidence from surface dyslexia. *Cortex* **74**, 191–205.
Liberman, Mark & Alan S. Prince (1977). On stress and linguistic rhythm. *LI* **8**, 249–336.
Lindsey, Geoff (2019). *English after RP: standard British pronunciation today*. Cham: Palgrave Macmillan.
Liu, Dungang & Heping Zhang (2018). Residuals and diagnostics for ordinal regression models: a surrogate approach. *Journal of the American Statistical Association* **113**, 845–854.
Lowenstamm, Jean (1999). The beginning of the word. In John R. Rennison & Klaus Kühnammer (eds.) *Phonologica 1996: syllables!?* The Hague: Thesus, 153–166.
McKinnon, Richard, Mark Allen & Lee Osterhout (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *NeuroReport* **14**, 883–886.
McMahon, April (2001). Review of Hammond (1999). *Phonology* **18**, 421–426.
Pastizzo, Matthew J. & Laurie B. Feldman (2004). Morphological processing: a comparison between free and bound stem facilitation. *Brain and Language* **90**, 31–39.
Pater, Joe (1994). Against the underlying specification of an exceptional English stress pattern. *Toronto Working Papers in Linguistics* **13**, 95–121.
Pater, Joe (1995). On the nonuniformity of weight-to-stress and stress preservation effects in English. Ms, McGill University. ROA #107.
Pater, Joe (2000). Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* **17**, 237–274.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press

Plag, Ingo & Sonia Ben Hedia (2018). The phonetics of newly derived words: testing the effect of morphological segmentability on affix duration. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.) *Expanding the lexicon: linguistic innovation, morphological productivity, and the role of discourse-related factors*. Berlin: Mouton de Gruyter, 93–116.

Pöchtrager, Markus (2022). The weak vowels of English: openness as structure. Paper presented at the 29th Manchester Phonology Meeting, online, May 2022.

Poldauf, Ivan (1984). *English word stress*. Oxford: Pergamon Press.

R Core Team (2023). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/.

Raffelsiefen, Renate (1999). Diagnostics for prosodic words revisited: the case of historically prefixed words in English. In T. Alan Hall & Ursula Kleinhenz (eds.) *Studies on the phonological word*. Amsterdam: Benjamins, 133–201.

Raffelsiefen, Renate (2007). Morphological word structure in English and Swedish: the evidence from prosody. In Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli & Sergio Scalise (eds.) *Proceedings of the fifth Mediterranean Morphology Meeting*. Bologna: University of Bologna, 209–268.

Rastle, Kathleen & Max Coltheart (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language* **42**, 342–364.

Ripley, Brian, Bill Venables, Douglas M. Bates, Kurt Hornik, Albrecht Gebhardt & David Firth (2019). MASS: support functions and datasets for Venables and Ripley's MASS. R package version 7.3-51.5. Available from https://CRAN.R-project.org/package=MASS.

Roach, Peter (2009). *English phonetics and phonology: a practical course*. 4th edition. Cambridge: Cambridge University Press.

Ross, John R. (1972). A reanalysis of English word stress. In Michael K. Brame (ed.) *Contributions to generative phonology*. Austin: University of Texas Press, 229–323.

Scheer, Tobias & Philippe Ségéral (2020). Elastic s+C and left-moving yod in the evolution from Latin to French. *Probus* **32**, 183–208.

Selkirk, Elisabeth O. (1980). The role of prosodic categories in English word stress. *LI* **11**, 563–605.

Selkirk, Elisabeth O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Shaol, Cyrus & Chris Westbury (2010). Neighborhood density measures for 57,153 English words. Database. Published online at https://www.psych.ualberta.ca/~westburylab/downloads/westburylab.arcs.ncounts.html.

Siegel, Dorothy C. (1974). *Topics in English morphology*. PhD dissertation, Massachusetts Institute of Technology.

Sonderegger, Morgan & Partha Niyogi (2013). Variation and change in English noun/verb pair stress: data and dynamical systems models. In Alan C.L. Yu (ed.) *Origins of sound change*. Oxford: Oxford University Press, 262–284.

Steriade, Donca (1997). Lexical Conservatism. In *Linguistics in the morning calm: selected papers from SICOL 1997*. Seoul: Hanshin Publishing Company, 157–179.

Steriade, Donca & Juliet Stanton (2020). *Productive pseudo-cyclicity and its significance*. Paper presented at LabPhon 17, University of British Columbia (online), July 2020.

Szigetvári, Péter (2017). Strengthening in unstressed position: we happy? In Péter Szigetvári (ed.) *70 snippets to mark Ádám Nádasdy's 70th birthday*. Budapest: Department of English Linguistics, Eötvös Loránd University. Published online at http://seas.elte.hu/nadasdy70/szigetvari.html.

Szigetvári, Péter (2018). Stressed schwa in English. *The Even Yearbook* **13**, 81–95.

Szigetvári, Péter (2020). Posttonic stress in English. In Krzysztof Jaskula (ed.) *Phonological and phonetic explorations*. Lublin: Wydawnictwo KUL, 163–189.

Szigetvári, Péter (2022). Unstressed vowels in English: distributions and consequences. *Acta Linguistica Academica* **69**, 4–16.

Taft, Marcus (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes* **9**, 271–294.

Taft, Marcus & Kenneth I. Forster (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* **14**, 638–647.

Taft, Marcus, Gail Hambly & Sachiko Kinoshita (1986). Visual and auditory recognition of prefixed words. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **38A**, 351–366.

Tokar, Alexander (2019). Pretonic alphabetic *o* in Present-Day English. *Language Sciences* **74**, 1–23.

Trevian, Ives (1993). *Phonographématique, phonologie et morphophonologie des consonnes en anglais contemporain*. PhD dissertation, Université Paris VII.

Upton, Clive & William A. Kretzschmar (2017). *The Routledge dictionary of pronunciation for current English*. London: Routledge.

Wells, John C. (1990). Syllabification and allophony. In Susan Ramsaran (ed.) *Studies in the pronunciation of English: a commemorative volume in honour of A.C. Gimson*. London: Routledge, 76–86.

Wells, John C. (2008). *Longman pronunciation dictionary*. 3rd edition. London: Longman.

Zhang, Yuhan (2021). Partial dependency of vowel reduction on stress shift: evidence from English *-ion* nominalization. In Ryan Bennett, Richard Bibbs, Mykel Loren Brinkerhoff, Max J. Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel & Maya Wax Cavallaro (eds.) *Proceedings of the 2020 Annual Meeting on Phonology*. Washington: Linguistic Society of America, 9 pp.

Zuur, Alain F., Elena N. Ieno & Chris S. Elphick (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* **1**, 3–14.

https://doi.org/10.1017/S0952675725100122 Published online by Cambridge University Press