



**HAL**  
open science

# Opaque morphology and phonology: Historical prefixes in English

Quentin Dabouis, Jean-Michel Fournier

► **To cite this version:**

Quentin Dabouis, Jean-Michel Fournier. Opaque morphology and phonology: Historical prefixes in English. *Journal of Linguistics*, In press, 10.1017/S002222672400015X . hal-04577381

**HAL Id: hal-04577381**

**<https://hal.science/hal-04577381>**

Submitted on 16 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Distributed under a Creative Commons Attribution 4.0 International License



RESEARCH ARTICLE

# Opaque morphology and phonology: Historical prefixes in English

Quentin Dabouis<sup>1</sup>  and Jean-Michel Fournier<sup>2</sup>

<sup>1</sup>Université Clermont Auvergne, Laboratoire de Recherche sur le Langage (UR 999)

<sup>2</sup>Université de Tours, Laboratoire Ligérien de Linguistique (UMR 7270)

**Corresponding author:** Jean-Michel Fournier; Email: [fournier@uca.fr](mailto:fournier@uca.fr)

**Received:** 11 January 2023; **Revised:** 06 April 2024; **Accepted:** 08 April 2024

**Keywords:** English; phonology–morphology interface; prefix; semantics

## Abstract

In most of the literature on English phonology, historically prefixed words such as *contain*, *respect* or *submit* are seen as having no morphological structure synchronically. However, such words were treated as complex in the early generative literature and are still analysed in that way in part of the literature. In this paper, we seek to review the evidence for the claim that such words are simplex words, which predicts that they should pattern with words with no internal structure in their phonological and morphological behaviours and in psycholinguistic experiments. We show that the evidence does not support that claim and shows that these words should be treated as morphologically complex units, although they differ from words with productive morphology. As these words tend to be partly or entirely opaque semantically, this raises the question of how such structures may be learned. We argue that the recurrence of forms is the main factor leading to their identification and lay out a possible order of acquisition of various morphological structures. Finally, we argue that theories of phonology may account for this by allowing the reference to morphological constituents whose semantics are impoverished.

## 1. Introduction

Opaque prefixed words have been the objects of specific analyses in Chomsky & Halle (1968), using the ‘=’ boundary but, after this boundary was rejected by Siegel (1974; 1980), most authors have progressively stopped referring to these units, and those words are often listed among morphologically simple words. In the literature on morphology, the status of historically prefixed verbs such as *contain*, *refuse* or *submit* in contemporary English is a well-known problem (Aronoff 1976: 55; Anderson 1992: 55; Plag 2003: 30–33; Katamba & Stonham 2006; Harley 2009; 2014; Bauer et al. 2013: 15–16; Mudgan 2015) because they contain recurring forms with no clearly identifiable meaning, which therefore constitutes a challenge to the standard definition of the morpheme as the minimal meaningful unit.<sup>1</sup>

<sup>1</sup> We will refer to those words as “opaque prefixed words” although they are not all opaque to the same degree (see §2.4).

The aim of this paper is not to discuss how the constituents of opaque prefixed words should be called (formatives, pseudo-morphemes, etc.) but to review different types of evidence that justify their identification as morphologically relevant units, and thus the distinction of those words from morphologically simple words. If opaque prefixed words function like morphologically simple words, then we should expect them to pattern with words with no identifiable structure: their phonological behaviour should resemble that of simplex words and psycholinguistic studies investigating them should report that speakers do not treat them like complex words. Others have developed similar views, relying partly on some of the evidence that we will review (Baeskow 2006, Harley 2009) but our focus is here on the phonological behaviour of such words.<sup>2</sup> Other kinds of evidence are put forward as corroborating evidence for the analysis that we propose. Another difference with those previous works is that they focus on prefixed words of Romance origin, whereas we include Germanic prefixes such as *be-* or *for-* in the analysis. Those have phonological properties that are comparable to Latinate prefixes (detailed in §2.1) but are less common and do not occur in as many forms as Latinate prefixes. The roots found in prefixed words of Germanic origin tend to be free, although there are some bound roots (e.g. *begin*, *believe*).

We will start by going through the different types of evidence (phonological, morphological and psycholinguistic) and will show that they strongly support the view that opaque prefixed words are still complex words in contemporary English (§2). Then, we will address the issue of learnability (§3) and the nature of the lexical representations of such words (§4).

Before we begin reviewing the evidence, let us note that not all studies deal exactly with the same type of words. Indeed, the words investigated in the studies reviewed in §2.1 sometimes include items in which the base is free but the following sections in §2 almost exclusively deal with items containing bound roots. As we will see, even among the latter words, there is some variability in the semantic contribution made by the prefix and/or the root which is generally overlooked.

## 2. Available Evidence Regarding Opaque Prefixed Words

### 2.1. Phonological evidence

#### 2.1.1. Vowel reduction: the initial pretonic position

Virtually all major works on the phonology of English recognize that ‘Latinate’ prefixes<sup>3</sup> have a specific reduction behaviour (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 2009; of the other3; Collie 2007: 129, 215, 318–319). Initial pretonic closed syllables normally do not contain reduced vowels (1a) whereas prefixes in that position almost always do (1b).

<sup>2</sup> Because of this focus on phonology, we are using a ‘piece-based’ approach to morphology throughout the paper, as this is the assumption adopted by most phonological theories. However, the facts that we lay down may certainly be captured in word-based models of morphology.

<sup>3</sup> This restriction to Latinate prefixes is most likely an overinterpretation of the fact that Chomsky & Halle (1968) never use examples containing Germanic prefixes. Guierre (1979) notes that Germanic prefixes have the same behaviour (e.g. *b[i~ə]spéak*, *b[i~ə]stów*, *b[i~ə]stride*).

- (1) a. /[a]mpóon    b. [ə]dvánce  
       p[ɒ]ntificate    c[ə]ndéense  
       t[ɛ]chnique      s[ə]btráct

In other environments, such prefixes may have unreduced vowels (e.g. [á]dverb, c[ɔ́]ncert, s[á]bscript), and we also find alternations between reduced (2a) and full vowels (2b) between morphologically related words.

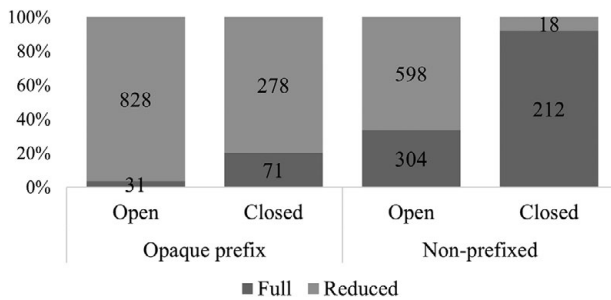
- (2) a. [ə]dvántage    b. [à]dvantágeous  
       c[ə]mpúte        c[ɔ́]mputácion  
       d[ə]báuch        d[è]bauchée

To our knowledge, the only proposed explanation for this particular behaviour is given by Hammond (2003). Building on previous work by Fidelholtz (1975), he argues that, since high-frequency items tend to reduce more than low-frequency items and because these prefixes are quite frequent, it is hardly surprising that these prefixes should be reduced. This is the only proposal that we are aware of in which the frequency of an affix (as opposed to that of a whole word) is assumed to be relevant, and it is still a hypothesis that has not been tested empirically. Formally, Hammond analyses the difference between high- and low-frequency items by assuming that they are indexed to different constraints (Pater 2000, 2009a). The effect of lexical frequency is usually interpreted in information-theoretic terms: more frequent items are more predictable and less informative, and so it is expected that they undergo more phonetic reduction, while less frequent and less predictable items are expected to undergo less reduction.<sup>4</sup>

Raffelsiefen (2007) proposes an alternative analysis. She claims that vowel reduction in closed initial syllables is specific to verbs. However, she contrasts only prefixed verbs (e.g. *contain*, *obsess*, *suspect*) and non-prefixed nouns (e.g. *canteen*, *pontoon*, *sestet*). There are only two disyllabic non-prefixed verbs with final primary stress and a closed initial syllable, *bl[a]sphéme* and *sh[a]mpóo*, but both have full vowels in their first syllable. However, they may be unreliable because of the existence of a related noun (*shampoo*) or a morphologically related word with stress on that syllable (*bláspheemy*). Likewise, prefixed nouns with final stress are scarce, but all have reduced vowels (e.g. *c[ə]nstráint*, *c[ə]ntéempt*, *s[ə]spéense*). Again, that behaviour can be attributed to the corresponding verbs (*constrain*, *contempt*, *suspend*). Therefore, one needs to go beyond those data to establish what the presence of reduced vowels should be attributed to. The following study that we discuss partly provides us with evidence that prefixation is indeed the key criterion.

Dabouis & Fournier (submitted), in a large-scale multifactorial study of vowel reduction in British English using dictionary data, report a significant difference between words containing a monosyllabic prefix (defined based on etymology – words with compositional semantics are excluded for that part of the study) and words which do not. Prefixation is tested in several regression analyses including multiple covariates (e.g. word frequency, syllable structure, spelling), and comes out as a highly significant predictor of vowel

<sup>4</sup> Such studies often deal with quite gradient kinds of reduction, while certain words are systematically realized with reduced vowels. Thus, the kind of ‘vowel reduction’ that we see in (1) and (2) has quite likely undergone phonologization, following the life cycle of phonological processes (Bermúdez-Otero 2015), and is now probably best described as a form of lexical redundancy rule and with a few lexical exceptions.



**Figure 1.** Vowel reduction in the initial pretonic syllable of non-derived words (figures from Dabouis & Fournier *submitted*).

reduction, independently of those covariates. Figure 1 reproduces their results for words which are not derived from a free base and whose initial pretonic syllable does not contain an orthographic digraph, which they found to impact vowel reduction.

As can be seen in Figure 1, non-prefixed words have reduced vowels in 66% of cases in which the syllable is open and 8% of cases in which it is closed, while prefixed words have 96% and 80% of reduced vowels in these environments, respectively. The difference persists in stress-shifted derivatives (e.g. *contéxtual* ← *cóntext*; *solidity* ← *sólid*), in open syllables (48% vs. 80%) and closed syllables (9% vs. 73%).<sup>5</sup> That last point clearly contradicts Raffelsiefen's (2007) claim above that the presence of reduced vowels is attributable to syntactic categories. She discusses disyllables exclusively, and we saw that it is virtually impossible to tease apart the effects of prefixation from those of syntactic categories in disyllables. The longer stress-shifted derivatives studied by Dabouis & Fournier, which are mainly nouns and adjectives, have reduced vowels significantly more often when they are prefixed than when they are not, which shows that those prefixes reduce regardless of syntactic categories.

To sum up, in this section, we saw that words which have semantically opaque prefixes in their initial pretonic position have reduced vowels significantly more often than non-prefixed words.

### 2.1.2. Vowel reduction: the second syllable of disyllabic words with initial primary stress

Using dictionary data, Guierre (1979: §4.2.6) observed that the presence of a reduced vowel in the second syllable of disyllabic non-prefixed words with initial primary stress appears to be the rule (87% of words; e.g. *álb[ə]rt*, *báll[ə]st*, *hárv[ɪ]st*, *hón[i]*, *sérv[ə]nt*, *pílf[ə]r*), while opaque prefixed words (which he calls 'inseparable') follow the opposite tendency (84% of words have unreduced vowel; e.g. *ádv[ɜ:]rt*, *cómm[ɛ]nt*, *cóntr[ɑ:]st*, *díg[ɛ]st*, *súrv[ei]*,

<sup>5</sup> An anonymous reviewer points out that one could argue that this statistical pattern is just a reflection of some historical generalization, not a part of the synchronic grammar. We pointed out in note 3 that we do not assume that we are here dealing with an across-the-board process of reduction but rather a sort of lexical redundancy rule, much like other processes such as Velar Softening or Trochaic Shortening. In any case, this is still synchronically a phonological characteristic of those prefixes (though not a systematic one) and noncword experiments would be needed to establish whether native speakers of English have it as part of their grammar. There is a large literature on the fact that speakers can learn probabilistic trends in their lexicon and apply them to new items ((Zuraw 2010, Becker et al. 2011, Moore-Cantwell 2016) and so it could be expected to extend to that pattern.

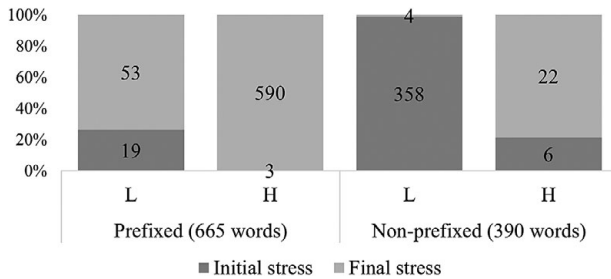
*tránsf*[3:ɹ]. However, this may be due to the fact that disyllabic nouns with trochaic stress which are related to a verb with final stress (e.g. *escort*, *import*, *project*) tend to have full vowels in their second syllable (Fudge 1984: 32, 167; Poldauf 1984: 38). Although the majority of words concerned by this generalization are prefixed, it occasionally extends to non-prefixed words (e.g. *fěrm*[ɛ]nt (n), *tórm*[ɛ]nt (n)). The only study to have sought to untangle both possible generalizations is Dahak (2011), which reports a study based on data from pronunciation dictionaries on 9500 disyllabic words with initial stress, 5.5% of which are prefixed. The overall figures that she reports are similar to those found by Guierre (1979), but she provides more detail on this issue: in her dataset, a quarter of words have a related form with final stress (e.g. *abstract*, *perfect*, *protest*) and 96.3% of those words are attested with a full vowel in their second syllable. In contrast, the remaining three quarters which do not have a related form with final stress (e.g. *abject*, *expert*, *transept*) may have a full vowel in 81.5% of cases. Therefore, even if we isolate the effects of the existence of a related word with primary stress on the second syllable, opaque prefixed words still have a behaviour that is the exact opposite of non-prefixed words. Although Dahak does not always provide word counts and does not conduct a multifactorial analysis, the magnitude of the difference between prefixed and non-prefixed words suggests that this difference would persist in a more controlled study.<sup>6</sup>

In this section, we saw that disyllabic opaque prefixed words with initial primary stress often have full vowels in their second syllable, unlike non-prefixed words which nearly never have full vowels in that position. In the case of nouns, we saw that this cannot systematically be attributed to the fact that such nouns are related to a verb with final primary stress (although it can further increase the chances of having a full vowel).

### 2.1.3. Primary stress

In a large dictionary study on primary stress placement in verbs, Dabouis & Fournier (2023) test two generalizations proposed in the literature regarding the parameters that determine the position of primary stress: the syllabic weight of the final syllable and the morphological structure of verbs. The ‘weight-based generalization’, which is inherited from Chomsky & Halle (1968) and has been preserved in most of the subsequent literature in various forms (see e.g. Hayes 1982; Burzio 1994: 43; Hammond 1999: 263), predicts that stress should be final if the final syllable is heavy, and penultimate otherwise. This follows from the assumption that stress is quantity-sensitive and rightmost in English and that the final consonant of verbs does not contribute weight. One noticeable exception to this is that trisyllabic verbs that have a heavy final syllable usually have antepenultimate stress, except if their final syllable is a root (Chomsky & Halle 1968, Liberman & Prince 1977, Hayes 1982). The ‘morphology-based generalization’ predicts that stress should fall on the root in opaque prefixed verbs. These parameters are evaluated on a subset of their data which

<sup>6</sup> The difference between putative prefixes and roots seen in these two sections is striking. We see two ways to explain it. First, if one assumes that these words are indeed composed of a prefix and a root, then the difference may be attributed to the different morphological nature of these elements: even though certain roots are bound, there remains a difference between prefixes and roots. As roots are the central part of the word, it is possible that there is a pressure to maintain a certain level of prominence on these elements, but not on prefixes. Another explanation may lie in frequencies: if prefixes are more frequent than roots, then they are more likely to reduce. Although we have not verified this, it is likely to be true as prefixes generally occur in more words than roots do.



**Figure 2.** Position of primary stress in disyllabic verbs depending on the weight of the final syllable (H = heavy and L = light) and on the presence or absence of an opaque prefix (from Dabouis & Fournier 2023).

contains both morphologically simple words and words containing bound roots, where the presence of affixes is determined using etymology. They find that both weight and prefixation are highly significant predictors of the position of primary stress in a binary logistic regression. Their results show that both parameters make identical predictions in the majority of cases, many of which are either prefixed with a heavy final syllable or non-prefixed with a light final syllable. The figures that they report for disyllabic words are reproduced in Figure 2. Remember that those do not include opaque prefixed words in which the base is free (e.g. *become*, *express*, *reclaim*), in which primary stress is base-initial in 99% of cases, and so the word treated as ‘prefixed’ in Figure 2 all have bound roots.

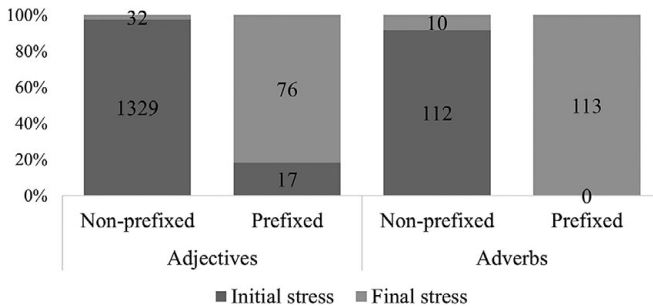
Overall, the two parameters have comparable efficiencies, but both fail to capture part of the data. The weight-based generalization fails to capture prefixed words with a light final syllable and final stress (e.g. *attack*, *commit*, *repeal*) whereas the morphology-based generalization fails to capture non-prefixed words with a heavy final syllable and with final stress (e.g. *lament*, *molest*, *usurp*). In longer verbs, the same observation holds, as primary stress is base-initial in 115/138 (83%) of cases (e.g. *apprehend*, *contradict*, *demolish*, *entertain*, *intermit*, *remember*, *rétrogress*, *surrender*). Crucially, trisyllabic verbs with two prefix syllables and a heavy final mostly have final stress, and this differs from non-prefixed verbs (e.g. *dámascene*, *gállivant*, *mánifest*). If the final is light, penultimate stress is more systematic for verbs with a monosyllabic prefix than in non-prefixed verbs (e.g. *sequéster* vs. *mónitor*). These results show that both generalizations constitute viable analyses of stress placement in verbs.

Moreover, they report that, in American English, there is a significant difference between words in <-ate> which contain a prefix (e.g. *ablate*, *debate*, *relate*) and those which do not (e.g. *castrate*, *dictate*, *rotate*), as ‘the former have final stress as their main pronunciation in 13/14 words, while the latter have final stress in 9/35 words’ (chi-squared=15.609,  $df=1$ ,  $p < .001$ ), as observed previously by Halle & Keyser (1971: 68–69) and Guierre (197: §1.2.2).

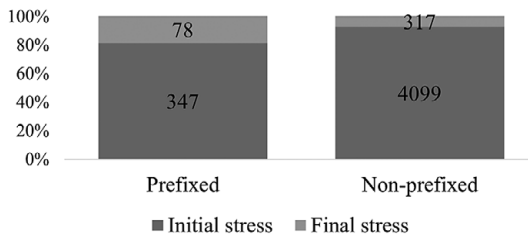
Guierre (1979) reports similar figures for verbs but also gives the figures shown in Figure 3 for disyllabic adjectives and adverbs.

There too, there is a highly significant difference between prefixed and non-prefixed words, both for adjectives (chi-squared = 786.02,  $df=1$ ,  $p < .001$ ) and adverbs (chi-squared = 194.53,  $df=1$ ,  $p < .001$ ). In longer words, Guierre (1979) reports that non-suffixed adjectives have antepenultimate stress if the final syllable is heavy, and that no difference is observed between prefixed (e.g. *ábsolute*, *dérelíct*, *óbsolate*, *résolute*) and non-prefixed words





**Figure 3.** Position of primary stress in disyllabic adjectives and adverbs depending on the presence or absence of an opaque prefix (figures from Guierre (1979))



**Figure 4.** Position of primary stress in disyllabic nouns depending on the presence or absence of an opaque prefix (figures from Guierre (1979)).

(e.g. *ádamant, fúribund, máritime, móribund*) but notes that those classes are very restricted (15 words in total). In non-suffixed adjectives that end in a light syllable (*/-VC/*), prefixed words have penultimate stress (e.g. *intrépid, insípíd, ímplicit, explícít, decrépít*) while non-prefixed words have antepenultimate stress (e.g. *fínikin, filemot, áliquot, Sáracén, mínikin, sínister*) but again the relevant classes are very small (the 11 examples that we gave are the complete list).<sup>7</sup>

The results reported by Guierre (1979: §4.2.5.5 and §4.2.6) also show a significant difference in how primary stress is assigned to disyllabic nouns (chi-squared = 63.119, *df* = 1, *p* < .001), though not as strong as that found in verbs, adjectives and adverbs, as shown in Figure 4.

As can be seen in Figure 3, he finds that disyllabic nouns have initial stress more often when they are not prefixed (4,099/4,416 – 93%; e.g. *beauty, radish, visa*) than when they contain an opaque prefix (347/425 – 81%; e.g. *absence, congress, digest*). If one considers the figures that he reports regarding how many words are prefixed (among the words that do not contain any productive morphology or stress-affecting endings), we can see that verbs are, in their vast majority, prefixed (1,032/1,345 – 77%), but the same is not true of nouns (425/4,841 – 9%).

<sup>7</sup> Guierre also notes another property of long prefixed adjectives: their penultimate vowel is always short, as in most disyllabic adjectives (e.g. *r[a]bid, t[i]mid, Sp[a]nish, c[i]vil, pr[v]per; c[ɛ]ver*). He makes the same observation in verbs (e.g. *cons[ɪ]der, del[ɪ]ver, dev[ɛ]lop, inh[ɪ]bit, b[v]ther; g[a]ther, h[ɛ]rald, h[v]ver, tr[a]vel*; see also Nessly (1982) and Dabouis (2018)). Crucially, this is different from nouns, in which penultimate long vowels are a lot more common.



For longer nouns, Guierre reports no difference between non-prefixed and prefixed words (e.g. *cómpany*, *déficit*, *éxodus*, *précipice*), but he does not provide figures or statistical tests to base his observation. Therefore, further research is needed to replicate and consolidate that finding. However, the results on disyllables are robust: there is a difference between prefixed and non-prefixed nouns, although the difference is considerably smaller than that observed in verbs, adjectives and adverbs, which are almost systematically stressed on their base. This often results in stress patterns that are distinct from those of non-prefixed words.

#### 2.1.4. The diachronic evolution of stress in verb–noun pairs

Sonderegger & Niyogi (2013) report a study of the diachronic evolution of disyllabic verb–noun pairs such as *convict*, *concrete* or *exile*. They find that most observed changes lead to diatonic pairs in which the verb has final stress and the noun has initial stress, although most pairs do not change if they are in what they call a ‘stable state’ (both are stressed on the first syllable, both are stressed on the second syllable, or the verb has final stress, and the noun has initial stress). They find that the words which share a prefix tend to evolve in a similar direction, and they note that that is ‘particularly interesting because it suggests that it is a shared morphological prefix rather than simply shared initial segments which correlates with trajectory similarity’. Therefore, one might interpret the fact that words that share a prefix evolve in a similar direction as an indication that those prefixes are detected by language users of English.

#### 2.1.5. Secondary stress

Robust effects of opaque prefixation were reported by Arndt-Lappe & Dabouis (submitted) in suffixal derivatives. The study sought, among other things, to replicate Collie’s (2007) finding that weak preservation may fail and that this failure can be attributed to the relative frequency of the base and its derivative. Weak preservation designates the preservation of the primary stress of a base as a secondary stress in its derivative (e.g. *oríginál* → *oríginálicity*). The earlier literature assumed that this phenomenon was categorical (Kiparsky 1979, Hammond 1989, Halle & Kenstowicz 1991) but Collie found that there were instances of (possible) preservation failure (e.g. *authórity* → *authòritárian* ~ *àuthoritárian*). Arndt-Lappe & Dabouis report two studies, one based on dictionary data and one based on oral data. While evidence of relative frequency was only found for the oral data, both studies report that opaque monosyllabic prefixes favour second-syllable stress preservation.<sup>8</sup>

Therefore, there is some evidence of the stress-repellent effect of opaque prefixes with regards to pretonic secondary stress placement.

#### 2.1.6. Phonotactics

Guierre (1990) proposed that certain consonant clusters which are illicit in simplex words can indicate the presence of a morphological boundary. He notes that the only clusters of two obstruents which are attested word-medially in simplex words are voiceless clusters.

<sup>8</sup> Note however, that this effect could not be teased apart from that of the relative weight of the first two syllables in the study of the oral data.

Hammond (1999: §3.3) has conducted a more extensive survey of distributional regularities in English and found slightly different results but also identifies medial clusters which can be found only (or mostly) in constructions with Latinate prefixes, such as:

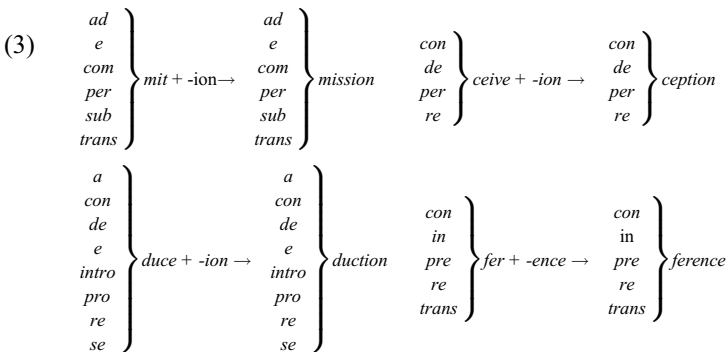
- those with voicing disagreements involving voiceless fricatives: [bs] *abcess*, *absence*, *subsidy*, [bf] *obfuscate*;<sup>9</sup>
- certain clusters containing a voiced stop followed by a sonorant: [bm] *submerse*, [dm] *admire*, [bn] *obnoxious*;
- most clusters containing a voiced stop followed by a voiced fricative or affricate: [bv] *obvious*, [bz] *absolve*, [dv] *advantage*, [bdʒ] *object*, [gdʒ] *suggest*;
- [dh] in *adhere*, a cluster normally only created by concatenation, e.g. *childhood*, *madhouse*.

Therefore, it can be argued that these clusters function as ‘boundary signals’ and can favour the recognition of a structure in these words. It is also possible that there are other phonotactic elements which may make words more or less likely to be interpreted as prefixed. For example, many roots end in VVC or VCC, with consonant clusters that are very typical of such roots (e.g. [pt, kt]).

Therefore, although more investigation is needed regarding which phonotactic cues might affect how likely it is for a word to be interpreted as prefixed, some word-internal clusters that do not occur in simplex words are found in opaque prefixed words and are therefore available to listeners as boundary signals marking the prefix–root boundary.

## 2.2. Morphological evidence: selectional restrictions and allomorphy

Aronoff (1976: 12–15) suggests that units such as *fer* or *mit* are ‘somehow relevant to the morphology’ because the verbs sharing a bound root have the same standard nominalizing suffix, sometimes accompanied by root allomorphy, as shown in (3).



Aronoff argues that prefix–root words in *mit* all contain the same root and not different homophones. He also observes a similar phenomenon for verbs which share the same irregular inflection even in the absence of semantic similarity (Aronoff 1976: 14), as can be seen in Table 1.

<sup>9</sup> As pointed out by an anonymous reviewer, it would be interesting to see how those clusters are actually realized phonetically considering that voiced and voiceless plosives may neutralize in preconsonantal position. They suggest that we may see longer durations of the preceding vowel when followed by underlyingly [+voice] [-voice] clusters as compared to [-voice] [-voice] clusters. Studies on this issue would be needed to have better understanding of the acoustic behaviour of such clusters.

**Table 1.** Irregular inflection in prefixed and non-prefixed words sharing a root

Infinitive	Simple past	Past participle
<i>come – become</i>	<i>came – became</i>	<i>come – become</i>
<i>get – beget – forget</i>	<i>got – begot – forgot</i>	<i>gotten – begotten – forgotten</i>
<i>give – forgive</i>	<i>gave – forgave</i>	<i>given – forgiven</i>
<i>stand – understand – withstand</i>	<i>stood – understood – withstood</i>	<i>stood – understood – withstood</i>
<i>take – mistake – partake – undertake</i>	<i>took – mistook – partook – undertook</i>	<i>taken – mistaken – partaken – undertaken</i>
<i>wake – awake</i>	<i>waked ~ woke – awaked ~ awoke</i>	<i>waked ~ woken – awaked ~ awoken</i>

However, Plag (2003: 26–27) rejects the argument that this is evidence for semantically-impoorished morphemes on the grounds that these facts ‘can equally well be described in purely phonetic terms’. To claim that the morphology needs to refer to the roots in these words, one would need cases in which the end of the verb would contain a phonetic string that is identical to a given root and that this string should not actually be a root. Some have proposed such examples as *vomit* → \**vomission*, but Plag rejects that example on the grounds that it is not in fact phonetically identical to the verbs with the root *mit* because it is stressed in these verbs but not in *vomit*. This leads Plag to conclude that units such as *mit* or *fer* should not be treated as morphemes and that ‘there is no compelling evidence so far that forces us to redefine the morpheme as a morphological unit that can be without meaning’.

What Plag does not mention is the difficulty of finding relevant examples to test that claim. There are hardly any verbs which are non-prefixed and have final stress (in an exhaustive dictionary study, Dabouis & Fournier (2023) find only 31 such verbs). Therefore, the argument suffers from a lack of support of compelling evidence and the alternative of a phonetically-driven affix selection process seems *ad hoc*. Although affixes have been shown to be selected based on the morphological structure of the base or the stress pattern of the base (e.g. Kiparsky 1982, Raffelsiefen 2004, Bauer et al. 2013), they have not, as far as we know, been shown to be sensitive to such complex structures as a whole syllable along with its prosodic specifications (e.g. [-'mit] or [-'fɜ:]). Before concluding that there is ‘no compelling evidence’, one would have to review more than a single process, as we have done in the previous sections.

Following Aronoff’s claim, it could be assumed the alternations in (3) can function as cues that roots are morphological units and that the words that contain them should be analysed as complex. However, in a segment-shifting study, Melinger (2001) reports results suggesting that the complexity of words containing alternating roots is more difficult to access than that of those containing non-alternating roots. The study consisted in presenting participants with a source word (e.g. *cooperate*, *coconut*) in which (pseudo)affixes are highlighted simultaneously with the presentation of the target word (e.g. *author*), to which they are asked to attach the highlighted section of the source word and pronounce a response word (e.g. *co-author*). The experiment contained words with free stems (e.g. *defrost*), bound roots (e.g. *deceive*; some of which had several forms) and simple words (e.g. *denial*) and it was expected that it would be easier to shift the highlighted section if the source word is morphologically complex. She finds that this is indeed the case, although the results regarding

items with bound roots are unclear as they are not significantly different from simple words and items containing free stems, but simple words are significantly different from items with free stems. Moreover, alternating roots were found to make the task more difficult, which she interprets as a sign that alternating roots ‘have weaker relationships to their morphological family members’ than non-alternating roots. However, her study suffers from limitations, such as the small number of words used and the heterogeneity of the alternations considered (she included words with common alternations such as *depar*[t] → *depar*[tʃ]*ure* alongside cases like *revolve* → *revolution*), and her results have not been replicated.

Therefore, we have seen that certain roots have a stable allomorphic behaviour in the different words that contain them, and some roots tend to associate with specific nominalizing suffixes. The results reported by Melinger suggest that it is possible that root allomorphy actually impedes the identification of morphological structure more than it facilitates it, but more evidence is required to confirm her observations.

## 2.3. Psycholinguistic evidence

### 2.3.1. Lexical decision

Several psycholinguistic studies have shown that prefixed words with bound bases (generally referred to as ‘bound stems’ in that literature) and which have obscure semantics are processed differently from words without any morphological structure in lexical access, and are often processed similarly to complex words with free bases (referred to as words with ‘free stems’). Taft & Forster (1975) (Experiment III; later replicated with more thoroughly controlled materials in Taft (1994)) found that, in a lexical decision task, participants took longer to reject prefixed non-words which contained a bound stem (e.g. *devade*, with *vade* from *evade*, *invade*) than those which did not (e.g. *deepoch*, with *poch* from *epoch*). Taft et al. (1986) also found that non-words with non-prefix initial elements (e.g. *tejoice*, *asiwhelm*, *lanlediate*, *asiwhast*), even if they contained bound stems, were rejected faster than those with initial prefixes (e.g. *dejoice*, *uniwhelm*, *dejoue*, *conlediate*). This suggests that stems are isolated via what the authors call a ‘prefix-stripping procedure’. A recent study, Coch et al. (2020), which compared word and non-word stimuli with bound stems (e.g. *discern*, *predict*, *disject*, *percern*) or with free morphemes (e.g. *cobweb*, *earring*, *cobline*, *bobweb*) with control monomorphemic stimuli (e.g. *garlic*, *minnow*, *gartus*, *buzlic*), found that stimuli with bound stems were processed differently in words and non-words. Indeed, they report no significant differences in accuracy and response times between bound stem and control monomorphemic words, but they find that bound stem and free morpheme non-words have comparable accuracies and response times, with lower accuracy and longer response times than monomorphemic non-words. They suggest that there is a ‘cost’ to the identification of morphological structure, even in the absence of semantics (in non-words), hence lower accuracies and longer response times.

To neutralize the effects of semantics, other researchers have conducted lexical decision tasks with masked<sup>10</sup> priming with prefixed words containing bound stems (e.g. *explore*) or

<sup>10</sup> There is variation in the evidence regarding the facilitation induced by bound stems in unmasked priming tasks. While certain studies report facilitation that is similar to free stems (Stanners et al. 1979) and independent from phonological similarity (Emmorey 1989), Marslen-Wilson et al. (1994) do not report any facilitation for bound stems. Pastizzo & Feldman (2004) assume that this may be due to the methodology used by the latter (i.e. cross-modal priming, which is known to be sensitive to semantics).

free stems (e.g. *distrust*) and found the same amount of priming facilitation for both structures, while controlling for orthographic factors (Forster & Azuma 2000, Pastizzo & Feldman 2004). Pastizzo & Feldman (2004) interpret this result as evidence that words composed of bound and free stems have similar representations.

Moreover, lexical decision latencies may be influenced by several factors such as the semantic transparency of the stem (as suggested by Taft & Kougious (2004); e.g. *venge* as in *revenge*, *avenge*, *vengeful*, *vengeance* has a clearer meaning than *ceive* in *receive*, *deceive*, *conceive*, *perceive*), the number of morphological relatives containing the same stem (Forster & Azuma 2000; Pastizzo & Feldman 2004; e.g. the stem *mit* can be found in more words than the stem *vive*) or the cumulated frequency of the stem in all the words in which it appears (Taft 1979; e.g. *proach* in *approach* and *reproach* has a higher frequency than *suade* in *persuade* and *dissuade*; lexical decision latencies are reported to be faster for *reproach* than for *dissuade* even though these words have equivalent printed frequencies).

In a review of the issue, Rastle & Davis (2008) discuss evidence from 19 studies on visual word processing and conclude that ‘morphological decomposition is a process that is applied to all morphologically structured stimuli, irrespective of their lexical, semantic or syntactic characteristics’. This morpho–orthographic decomposition would appear to take place in the early stages of the recognition process, independently of semantics (see Marslen-Wilson et al. (2008) and references therein). Although some studies do find a difference between transparent and opaque items, most of them assume that there is such a thing as morpho–orthographic decomposition which can apply to opaque items (e.g. Diependaele et al. 2009, Feldman et al. 2009; see Heyer & Kornishova 2018 for an overview). In a recent review of the issue, Beyersmann & Grainger (2023) draw similar conclusions, and they point out that semantically-blind decomposition is acquired late and that ‘morphological processing is primarily guided by semantics during the initial stages of reading acquisition’.

While most of the evidence comes from visual word processing, hence the term ‘morpho–orthographic’ used here (and in the literature), the little available evidence on spoken-word processing also suggests that morphological information is processed irrespective of semantic transparency in the first moments of processing (see Creemers (2023) for a review of spoken-word morphological processing and Emmorey (1989) for evidence on the auditory processing of prefixed words).

### 2.3.2. Reading studies

Four studies have shown a significant impact of prefixes on stress placement in reading tasks. Previous work had shown that words with segmental exceptions were read more slowly than regular words. For example, orthographic vowels followed by two consonants are usually pronounced as short vowels (e.g. *mint* [ˈmɪnt]), and so words which do not follow this generalization (e.g. *pint* [ˈpaɪnt]) tend to be read slower. Rastle & Coltheart’s (2000) aim was to see whether this would be true of stress exceptions as well: are exceptions to stress rules read slower than words that obey them? In a first reading experiment, where they classified disyllabic English words as ‘regular’ if they had initial stress and ‘irregular’ if they had final stress, no effect of stress regularity was found. The authors then redefined their model of stress regularity. Following Fudge (1984), they classified as ‘regular’ words with final stress which either contained a prefix (e.g. *ad-*, *de-* *sub-*) or a stress-taking suffix (e.g. *-ette*, *-oon*, *-ade*). In a second experiment, they asked participants to read 210 non-words, which could have either final stress (if they were prefixed or contained a stress-taking suffix)

or initial stress (otherwise). They compared the participants' productions to those of an algorithm programmed to predict stress based on the morphology of words and found a very high similarity between them. This led them to conclude that:

there seem to be a class of item that reliably takes second syllable stress. We suggest that this pattern of stress assignment may be related to the presence of morpheme-like orthographic segments—prefixes and stress-taking suffixes—which serve to place stress.

(Rastle & Coltheart 2000: 352)

Finally, they reproduced the first experiment with this new definition of regularity using 60 target words. They found significant effects of word frequency (there were longer naming latencies and higher error rates for low-frequency items) and stress irregular items (there were longer naming latencies and higher error rates for irregular items). They also found that 90% of errors were pure stress regularizations. Let us note that not all the possible determiners were taken into consideration in that study and that this could bias the results that they report.

The second study is that of Ktori et al. (2016), which focused on the reading of disyllabic prefixed words, which were treated as regular if they had final stress and irregular if they had initial stress, by patients with acquired surface dyslexia, which they define as 'an acquired disorder of reading in which the reading aloud of irregular words is impaired while the reading aloud of nonwords is spared'. Following Rastle & Coltheart (2000), they predicted that patients would regularize irregular prefixed words and assign them final stress. Five patients were asked to read 150 words, which fell into one of three categories: regular prefixed (e.g. *remind*), irregular prefixed (e.g. *reflex*) and non-prefixed with initial stress (e.g. *climate*). The results showed that patients were more likely to pronounce irregular prefixed words with final stress than non-prefixed words, therefore confirming the authors' hypothesis that 'prefixes repel stress'. As for the previous study, not all possible predictors of stress are controlled for (e.g. syntactic categories).

Ktori et al. (2018) investigate a larger range of potential determiners of stress placement in reading disyllables, some of which were not taken into consideration by previous studies. They conduct factorial experiments in which participants have to read non-words to disentangle potential cues to stress placement. These cues can be 'sublexical' (vowel length, orthographic weight, prefixation) or 'sentence-level' cues (grammatical category, rhythmic context). They report an effect of prefixation that is independent of other cues, which were all found to have an impact on stress placement in non-words.

The last study, Treiman et al. (2020), investigates several factors which have been claimed to influence how stress is placed in disyllabic English words and non-words, using experimental data (Studies 1–4) and dictionary data (Study 5). They report significant effects of the number of initial consonants, the number of final consonants, the syntactic context (nominal or verbal) and the presence of a prefix-like sequence (based on the classification used in Rastle & Coltheart (2000)), both for production data (sentence reading tasks) and in a relevant subset of the Unisyn lexicon (3,071 words).<sup>11</sup> Their results show that, in both cases, stress is highly predictable, although no factor has a 'none-or-all' effect, as could be expected

<sup>11</sup> Mousikou et al. (2017) shed doubt on the role of prefixes in stress assignment in disyllables, as prefixation "is unavoidably confounded with other potentially important cues to stress". Treiman et al.'s (2020) results confirm an independent effect of prefixation, while using what they show to be a better measure of orthographic weight.

by Rastle & Coltheart's (2000) model. The presence of a prefix is a factor which, among others, is probabilistically associated with a position of stress (second syllable), as shown by the mixed-models analyses that they report.

To sum up, those four studies all find prefixation to be a significant predictor of the position of stress in the reading of disyllabic words, although readers do not seem to apply a categorical rule not to assign stress to prefixes. The last two studies are the only ones to include syntactic categories among the covariates. This seems to be necessary given the important differences found between nouns and other syntactic categories that we reviewed in §2.1.3.

### 2.3.3. ERP

McKinnon et al. (2003) conducted an ERP study using 60 words containing bound stems (e.g. *submit*, *receive*), 60 non-words with the same prefixes and roots (*promit*, *exceive*), 60 control morphologically complex words (e.g. *bookmark*, *muffler*) and 60 control non-words with no apparent morphemes (e.g. *moobkark*, *flermuf*).<sup>12</sup> Participants saw the (non-) words on a screen and had to say whether or not they were English words. The N400 component of the ERP is known to be sensitive to lexical status (word vs. non-word) and frequency of the letter string: there is a higher amplitude for non-words and less frequent words than for words and specifically for more frequent words. They summarize the logic of their experiment as follows:

If readers treat the (experimental non-words (e.g. *promit*, *exceive*)) as unanalyzed wholes, then the non-words should elicit larger amplitude N400s than the real words. This follows from the fact that whereas the words are relatively frequent in English, the bound-stem non-words (taken as a whole unit) are non-words with zero frequency. Conversely, if readers decompose the words and non-words into their constituent morphemes, and if these morphemes are represented in the mental lexicon, then the words and non-words might elicit similar amplitude N400s; both the words and non-words are made up of morphemes with the same frequency.

(McKinnon et al. 2003: 883)

Their results show that only the control non-words elicited a larger N400 amplitude than the words. They interpret these results as being 'consistent with a model in which a morphologically complex letter string is parsed into its morpheme components, even when these morphemes are non-productive and semantically impoverished'.

A similar study, Coch et al. (2013), using a more controlled methodology and more participants (16, while McKinnon et al. (2003) had 12), found differences in N400 responses between items with free morphemes and those with bound morphemes. They interpret these results by saying that the N400 may actually correspond to a process of recomposition

<sup>12</sup> Those nonwords "were formed by permuting the letters of each control real-word into a pronounceable non-word with no apparent morphemes". An anonymous reviewer asks if those words were controlled for transitional probabilities. No mention of this is made in McKinnon et al. (2003), and one of the only two nonwords that are given in the paper, *moobkark*, contains an <bk> sequence that is hardly ever found in English (a search in Wells (2008) returns only three results: *Abkhazia*, *knobkerrie*, *lambkin*). It is possible that, if many of those stimuli have such low-frequency consonant clusters, they may have been interpreted as morphologically complex and so have biased the results.



guided by semantics, and that semantically-blind decomposition takes place earlier. They also report behavioural results (accuracy and reaction times) that are consistent with the literature reported in §2.3.1: items with bound or free morphemes differ from control items but not necessarily from each other.

Therefore, those two studies report inconsistent results regarding the N400 component of the ERP as to its sensitivity to meaning, so further replications of those studies would be welcome.

## 2.4. Semantic relics

In the previous sections, most of the work that we have cited treated words such as *conceive*, *emit* or *subtract* as semantically opaque, in the sense that their internal constituents have no clear identifiable meaning (although some use the term ‘semantically impoverished’). As we have seen, this is a central element for rejecting morphemic status to the constituents of those words, at least for those assuming that morphemes are defined at least partly by their meaning.

However, there have been proposals arguing that certain roots have some meaning that is shared by all the words that contain them. Fournier (1996) argues that this can be seen quite clearly with cases such as those in (4), in which the root *spect* is systematically associated with an idea of sight and the root *rupt* with an idea of breaking.

- (4) a. *aspect, circumspect, inspect, introspect, retrospect, suspect* (+ *spectacle*)  
 b. *abrupt, corrupt, disrupt, erupt, interrupt* (+ *bankrupt, rupture*)

This observation can be extended to prefixes, which sometimes contribute meaning, even though the meaning of the whole word is far from compositional. This can be seen in cases such as (5a), in which productive prefixes are attached to bound roots, or in (5b–c), in which the prefixes *re-* and *ex-* have a recurrent meaning, ‘back’ and ‘out,’ respectively, but that meaning is not the one found in the productive prefixes *re-* and *ex-*, ‘again’ and ‘former’. Thus, those are probably best analysed as homographic prefixes, some of which may only attach to bound bases.

- (5) a. *acephalous, decrease, include, intersperse, reflate, submerge*  
 b. *rebound* ‘to bound or spring **back** from force of impact’  
    *reflect* ‘to cast **back** (light, heat, sound, etc.) from a surface’  
    *repel* ‘to drive or force **back** (an assailant, invader, etc.)’  
 c. *exclaim* ‘to cry **out** or speak suddenly and vehemently’  
    *exhume* ‘to dig **out** of the earth’  
    *expel* ‘to drive or force **out** or away; discharge or eject’

A similar argument is developed by Baeskow (2006) within the framework of the feature-based theory of word formation (Baeskow 2002, 2004). Following Dowty (1991), Baeskow assumes that thematic roles are best seen as ‘cluster concepts rather than discrete categories’. Traditional thematic roles (e.g. patient, agent, instrument) are replaced by two proto-roles, which are defined over ‘lexical entailments’ (e.g. sentience, change of state, movement). Some of these entailments are associated with the ‘proto-agent’ and some with the ‘proto-patient’, and the number of entailments of each proto-role determines whether an argument is realized as a subject or a direct object. Baeskow claims that,

even from a synchronic point of view, each bound verbal root of Latin origin which is recurrent makes at least a small contribution to the semantic interpretation of the sets of verbs it heads in that it entails fragmental – or, borrowing Lieber’s (2004) terminology – skeletal thematic information. (Baeskow 2006: 19–20)

She defends this claim by focusing on six roots, *duce*, *scribe*, *ceive*, *sume*, *mit* and *port*, and she shows that they have such skeletal thematic features. We reproduce in (6) the argument structure that she claims is associated with those six roots, in which <E> refers to a referential argument and the subscripts refer to the entailments associated with the two proto-roles.

- (6) a. *duce* <E<p-agent<sub>[causation]</sub>, p-patient<sub>[causally affected]</sub>>>  
 b. *scribe* <E<p-agent<sub>[impulse, control, sentence, affection]</sub>, p-patient<sub>[affected]</sub>>>  
 c. *ceive* <E<p-agent<sub>[sentence]</sub>, p-patient<sub>[affected]</sub>>>  
 d. *sume* <E<p-agent<sub>[sentence, control]</sub>, p-patient<sub>[affected]</sub>>>  
 e. *mit* <E<p-agent<sub>[affection]</sub>, p-patient<sub>[change of state, affected]</sub>>>  
 f. *port* <E<p-agent<sub>[transaction]</sub>, p-patient<sub>[affected, change of state]</sub>>>

Baeskow claims that those roots have lexical entries which contain their argument structure and that the lexical entries of prefixed verbs ‘only add those properties which are not predictable from the entry of the root’. She also claims that the same type of ‘skeleton’ could be found for all recurrent Romance roots (although this is not demonstrated) and that her analysis predicts that those roots may be used to coin new words and provides support for her claim by citing examples from online sources (e.g. *subceive*, *exduce*, *rescribe*).

Let us note that Baeskow funds her analysis on occurrences in natural speech taken from online corpora, but that the analysis is qualitative. No quantification or statistical analysis is provided, and we are not told whether those entailments are different from what is found in non-prefixed verbs (e.g. is the feature [affected] absent in many dynamic verbs?). Therefore, the extension of her analysis to more roots and with quantitative aspects remains to be conducted.

In this section, we have seen that, at least for some prefixes and roots, there are semantic features that are shared by several or all the words that contain them, that they are not all opaque to the same degree, and those shared semantic features may participate in the identification of those constituents.

## 2.5. Interim summary

Let us sum up the main results that we have reviewed so far. We saw that opaque prefixed words differ from non-prefixed words in that:

- They near–systematically have reduced vowels in initial pretonic syllables;
- They quite generally have full vowels in the second syllable of disyllabic nouns with initial primary stress;
- They have primary stress on their root in verbs, adjectives and adverbs, which generates patterns which diverge from non-prefixed words;
- They have higher rates of final primary stress in nouns;
- They have undergone similar changes regarding the position of stress in verb–noun pairs if they share a prefix;

- They favour second syllable stress preservation in stress-shifting derivatives;
- They display specific phonotactic patterns, especially word-internal consonant clusters.

We also saw that reference to morphological complexity may be necessary to account for certain allomorphic behaviours and affix selectional restrictions. We saw that there is evidence from most psycholinguistic studies that we are aware of that prefixed words with ‘bound stems’ and opaque semantics are treated like complex words by participants, the only exceptions being experimental designs which do not neutralize semantic effects. Given the semantic opacity of such words, that result is actually to be expected. Finally, we saw that the prefixes and roots found in such words are not all entirely opaque and that roots such as *duce* or *mit* may be argued to share ‘skeletal’ thematic features.

Therefore, the evidence appears sufficient to reject the idea that opaque prefixed words are to be treated as simplex words. Any model seeking to maintain that idea would have to propose alternative accounts of the phonological processes and psycholinguistic results which we have reviewed. However, what we have seen so far raises several questions. First, one may wonder what the advantages of storing a word such as *permit* as *per+mit* could possibly be for the linguistic system. This question is discussed by Forster & Azuma (2000), who propose two answers. The first is that morphological structure is necessary for syllabification, independently of semantic transparency, as argued by Levelt et al. (1999). Levelt et al. discuss several studies on Dutch in which semantically empty formatives appeared to be necessary for syllabification, and so they ‘suggest that morphemes may be planning units in producing complex words without making a semantic contribution’. The second is that ‘the language acquisition system looks for and uses structure wherever it can be found’. A second question concerns the redundancy in lexical storage necessary if opaque prefixed words are analysed as complex words: as their meaning cannot be derived from that of their constituents, both the constituents and whole-word combinations may have to be stored in the lexicon. McKinnon et al. (2003) actually discuss this issue of redundancy and argue that a redundant system with both root and word storage ‘has considerable empirical and theoretical support’.

Finally, some may object that semantically opaque morphological structures are not learnable. We argue that the facts that we have reviewed can be accounted for by the assumption that speakers learn that opaque prefixed words are morphologically complex. As it cannot be assumed that they have access to etymological information, there must be synchronic cues which they can use. The phonological and morphological evidence reviewed in §2.1 and §2.2 may certainly constitute part of those cues, but other cues may also be used. In the following section, we argue that a key component is the distributional recurrence of prefixes and roots, sometimes facilitated by formal and/or semantic similarity with productive affixes or free stems.

### 3. Learning to Detect Complexity in Opaque Prefixed Words

The literature on the acquisition of derivational morphology shows that children first acquire morphological processes which involve little or no change in form or meaning between base and derivative (Tyler & Nagy 1989, Clark 2014, Fejzo et al. 2018). Frequency plays a role but it seems to be moderated by communicative usefulness (e.g. referential words are learned before function words, possibly because they are more useful to the child), more useful words being learned more quickly, even if they are not the most frequent, but frequency

creates more opportunities for lexical learning (Ravid et al. 2020). Thus, children generally start by using conversion and very productive suffixes such as the diminutive *-ie/y*, adjectival *-y* or the agentive *-er*, in which there is an important overlap in form and meaning between bases and derivatives. Prefixes appear to be more complex and appear later, starting with *un-*. Suffixes that are associated with stress and vowel differences between the bases and their derivatives (e.g. *-ion*, *-ity*, *-ic*) are acquired after children enter elementary school. It appears that semantic transparency is crucial in the first steps of the acquisition of derivational morphology. Although we cannot diagnose the acquisition of the ability to detect morphologically opaque structures through the production of novel forms (as is done for productive morphology), it is quite likely that such words are identified as complex quite late. Elements pointing in that direction come from recent studies on morpho–orthographic segmentation, which show that the process of semantically blind automatic segmentation of all (apparently) morphologically structured stimuli in the early stages of visual word recognition emerges late (Rastle 2019, Marelli et al. 2020, Beyersmann & Grainger 2023), probably during adolescence (Dawson et al. 2018).

Although morphologists may disagree regarding the way that morphological knowledge is represented, they usually assume that this knowledge is acquired through the comparison of sets of words. For example, Booij (2016) assumes that ‘language users can assign internal structure to a word if there is a systematic correlation between its form and meaning, based on the comparison of two sets of words’ and he then illustrates that mechanism using words that contain the agentive suffix *-er* and the corresponding verbs (e.g. *dancer* – *dance*; *fighter* – *fight*; *singer* – *sing*). As we have just mentioned, children can learn to identify bases that vary in form, and so we hypothesize that they can also learn to detect affixes that have little or no identifiable meaning, given sufficient distributional evidence. Taft (1994) assumes that a certain letter sequence does not need to be already identified as a prefix for a word to be analysed as having a prefix–root structure, this complexity ‘would simply emerge from the relationships that exist between words which share the same letter sequences’.<sup>13</sup> Forster & Azuma (2000) and Taft (1994) also suggest that the commonality of forms like *submit*, *admit*, *permit*, *prevent*, *advent*, *circumvent*, etc. would be sufficient to create subunits like *mit* and *vent*. Wurm (1997, 2000) also reports that ‘prefix likelihood’ (i.e. ‘the proportion of tokens beginning with a given letter string that are prefixed’, a measure introduced by Laudanna et al. (1994)) and frequency could impact the processing of words. Finally, Fournier (1996) assumes that the mechanism of ‘commutation’ (i.e. the mechanism inherited from European structuralism which allows for the identification of linguistic units when applied to minimal pairs, morphemes and phonemes, and is probably inherent to the speakers’ grammar) is the main cue which can be used by speakers to parse complex opaque structures.

Indeed, many prefixes and roots occur in several words, forming a network of prefix–root words, a sample of which can be seen in Table 2.<sup>14</sup>

<sup>13</sup> Although he reasons with written forms, the argument still holds for phonological forms.

<sup>14</sup> Some of those prefixes have forms that differ from that of the underlying prefix, for example as an effect of (historical) assimilation (as in the case of *ob-* and *sub-* in *offer* and *suffer*, respectively). While this is not necessarily an issue in itself for prefix identification (there are productive prefixes that have different forms; e.g. *in-* [m-, m-, n-, l-, u-]), it may complicate this identification, as argued later in this section. However, such changes do not usually affect the root and so the identification of word-internal structure may be facilitated by the identification of the root.

**Table 2.** An illustration of the network of prefixes and bound roots

Root Prefix	fect	fer	ject	mit	pose	sist	spire	tain	vert
a-/ab-/ad-	<i>affect</i>		<i>abject</i>	<i>admit</i>	<i>appose</i>	<i>assist</i>	<i>aspire</i>	<i>attain</i>	<i>a(d)vert</i>
con-/com-	<i>confect</i>	<i>confer</i>		<i>commit</i>	<i>compose</i>	<i>consist</i>	<i>conspire</i>	<i>contain</i>	<i>convert</i>
de-	<i>defect</i>	<i>defer</i>	<i>deject</i>	<i>demit</i>	<i>depose</i>	<i>desist</i>		<i>detain</i>	
e(x)-	<i>effect</i>		<i>eject</i>	<i>emit</i>	<i>expose</i>	<i>exist</i>	<i>expire</i>		<i>evert</i>
in-	<i>infect</i>	<i>infer</i>	<i>inject</i>		<i>impose</i>	<i>insist</i>	<i>inspire</i>		<i>invert</i>
ob-		<i>offer</i>	<i>object</i>	<i>omit</i>	<i>oppose</i>			<i>obtain</i>	<i>obvert</i>
per-	<i>perfect</i>			<i>permit</i>		<i>persist</i>	<i>perspire</i>	<i>pertain</i>	<i>pervert</i>
pre-	<i>prefect</i>	<i>prefer</i>							
pro-		<i>profer</i>	<i>project</i>		<i>propose</i>				
re-	<i>refect</i>	<i>refer</i>	<i>reject</i>	<i>remit</i>	<i>repose</i>	<i>resist</i>	<i>respire</i>	<i>retain</i>	<i>revert</i>
sub-		<i>suffer</i>	<i>subject</i>	<i>submit</i>	<i>suppose</i>	<i>subsist</i>	<i>suspire</i>	<i>sustain</i>	<i>subvert</i>

If the similarities allowing complex structures to emerge can be purely based on form, then it is to be expected that such structures will also be identified in entirely non-productive prefixes such as *se-* (a prefix which comes from Classical Latin *sē-*, where it was already archaic; Gaffiot 1934: 1409), which is found with different roots (7).

- (7) *secede* *accede, concede, precede, recede*  
*seclude* *conclude, exclude, include, occlude, preclude, reclude*  
*secrete* *accrete, concrete, discrete, excrete*  
*seduce* *adduce, conduce, deduce, educe, induce, produce, reduce, traduce*  
*segregate* *aggregate, congregate*  
*select* *collect, elect, intellect, neglect, recollect*

Psycholinguistic evidence can also inform us of some of the ways that this network might impact the emergence of complex structures: as was seen in §2.3.1, the recurrence of the root and its cumulated frequency have been found to affect lexical decision latencies.

Some of the roots found in prefixed words can also be found in other types of morphologically complex words. This can likely contribute to the likelihood that words containing these roots will be identified as complex, even more so when there can be semantic similarities of the kind discussed in §2.4, as found in the examples in (8).

- (8) *habit* cp. *habitat, habitation* ‘living’  
*spect* cp. *spectate, spectacle* ‘looking, sight’  
*rupt* cp. *rupture, bankrupt* ‘breaking’  
*venge* cp. *vengeance, vengeful* ‘avenging’  
*vive* cp. *vivace, vivid, vivify* ‘life’

If semantic transparency helps in the detection of word-internal structure, then words such as *aloud* or *surround* are probably analysed as morphologically complex before words such as *attain* or *submit*. Likewise, in words with bound roots where prefixes contribute semantics that are the same as those found in words with transparent semantics, the

identification of internal complexity is probably quite likely.<sup>15</sup> This is probably the case in isolated constructions (see (5a) and additional examples in (9)), in whole classes sharing semantics that may be associated with the prefix (see (5b)), and in series of words whose semantic opposition is based on that of their prefixes (10), such as the directional opposition in pairs in *ex-* and *in-*.

- (9) *circumscribe* ‘to draw a line **round**; to encompass; to **encircle**’  
*cohabit* ‘to live **together** as if married’  
*detoxify* ‘to **deprive** of poisonous qualities’  
*recapitulate* ‘to go through or repeat **again**; to summarize’
- (10) *deflate* ↔ *inflate* ↔ *reflate*  
*exhale* ↔ *inhale*  
*export* ↔ *import* ↔ *deport*

In those cases, the learner would operate comparisons between forms such as (9) and (10) and forms in which the same prefixes are attached to free bases and posit an internal structure that is similar to that of those words, except that the root is bound. Having analysed words such as (9) and (10) as internally complex could, in turn, facilitate the detection of complexity in words sharing the same roots, but with more opaque prefixes (11). For example, if one identifies *decelerate* as containing the negative prefix *de-*, then the bound base *celerate* will also be identified, and so *accelerate* will be more likely to be analysed as prefixed, even though its prefix is opaque. It is also possible that the meaning of the prefix of the second word in each of the pairs in (11), on account of the opposition with the other member, may be attributed semantic features that are the opposite of those of the prefix in the first member of the pair. For instance, *decelerate* contains the privative meaning usually found with the productive prefix *de-* (Bauer et al. 2013: Ch. 17), and so the prefix *ac-* in *accelerate* may be associated with a ‘positive’ meaning, in opposition to *de-*.

- (8) *decelerate* *cp.* *accelerate*  
*decrease* *cp.* *increase*  
*demote* *cp.* *promote*  
*dissimilate* *cp.* *assimilate*  
*subjacent* *cp.* *adjacent*

This would be very similar to the analysis put forward by Taft (2003) regarding how language users may assign meaning to bound constituents such as *hench*, which is only found in *henchman*, i.e. in association with a free morpheme:

[A] henchman is a man and, therefore, *hench* might be taken to refer to “the thing that a man does in order to be a henchman”. In this way, it might be possible to work out what a *henchwoman* or *henchdog* means if those words were coined.

(Taft 2003: 118)

<sup>15</sup> One reason for assuming this is that these words often have a phonological behaviour comparable to that of constructions with free stems (e.g. *deconstruct*, *re-do*, *unknown*; see Raffelsiefen (1993, 2007, 2015), Fournier (1996), Dabouis (2018, 2020)).

It is also possible that mere similarity of form facilitates the emergence of relevant units. One element in support of that view is that Dabouis (2024) reports that historically complex English proper names have phonological behaviours that resemble morphologically complex words more often when their first constituent resembles an existing word (e.g. *Davidson*, *Holywell*, *Maryland*), although there is no semantic connection to that word, than when it does not (e.g. *Corydon*, *Hutcheson*, *Sheringham*). This may manifest in prefixes and roots. In prefixes, the existence of homographic productive prefixes (e.g. *de-*, *pre-*, *pro-*, *re-*) may possibly favour complex interpretations.

Likewise, for roots, the existence of free forms possibly without any semantic relatedness (homophones and/or homographs; e.g. *beget*, *perform*, *understand*) could facilitate the detection of morphological complexity. Further studies will have to establish whether there is a difference in how speakers treat words for which there is a relatively clear semantic connection with the free form (e.g. *alight*, *oppress* or adverbs like *aloud*, *below*) and those for which that connection is obscure or absent (e.g. *inform*, *express*).

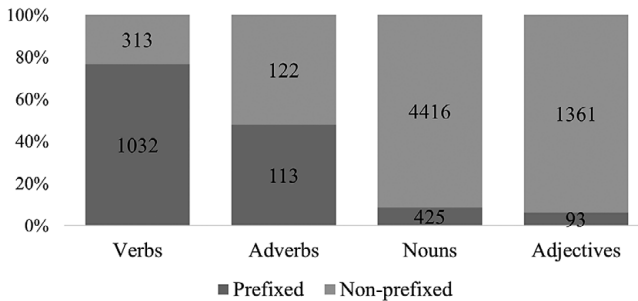
If inconsistency in form impedes the identification of morphological constituents, as reported by the literature on the acquisition of derivational morphology cited above and Melinger's (2001) study (see §2.2), then the prefixes and roots that have several realizations (e.g. *ad-/ab-/ac-/a-*; *con-/com-/cor-*; *ob-*, *o-*, *op-*; *tain/ten(tion)*; *mit/mission*; *veal/vel(ation)*) will probably be detected later than those which have a stable form (e.g. *de-*, *pre-*, *pro-*, *port*; *vest*). Moreover, among prefixes and roots that have different forms, all may not be equal, as alternations that occur relatively regularly in English, such as [t] ~ [ʃ] after *-ion* or *-ial* (e.g. *protec*[t] ~ *protec*[ʃ]*ion*) or vowel alternations caused by destressing (e.g. *decl*[é:]*re* ~ *dècl*[ə]*ration*) may complicate the identification of roots less than entirely unpredictable alternations (e.g. *receive* ~ *reception*). Therefore, we would assume that the detection of morphological structure in the various types of prefixed words that we have reviewed should proceed in an order that is after all quite trivial: units with the most constant overlap in form and meaning are detected first, and then the learner proceeds to detect more complex units displaying either variability in form and/or variable or unclear meaning.

As pointed out by an anonymous reviewer, we cannot exclude the possibility that different syntactic categories are treated differently, i.e. that opaque prefixation is detected more easily in certain categories than others.<sup>16</sup> If that is indeed the case, we would assume that the categories which contain the highest proportion of historically prefixed words are the ones for which prefixation will be detected first. If we use the figures given by Guierre (1979) for disyllables shown in Figure 5, then verbs would be first (77% are prefixed<sup>17</sup>), then adverbs (48%), then nouns (7%) and then adjectives (6%). Note that prefixed words with compositional semantics (which Guierre calls 'separable' prefixed words) are excluded from those counts.

<sup>16</sup> See Fournier (1996) for this very suggestion.

<sup>17</sup> Minkova & Zhou (2022) report that prefixation has been playing a role in the stress of English verbs since Old English, and that "there has been consistent lexical pressure to model verbs as a sublexicon with noninitial stress" with the massive borrowing of prefixed verbs during the Middle English period. They find that the proportion of verbs that are prefixed has gone from about 30% in 1200 to about 55% in 1500. This is consistent with Guierre's observation for contemporary English.





**Figure 5.** Proportions of prefixed words among disyllabic verbs, adverbs, nouns and adjectives according to Guierre (1979).

Finally, the phonological and morphological generalizations applying to opaque prefixed words, as we have suggested, resemble the sorts of semiproductive generalizations that are posited to occur at the stem-level in Stratal Phonology. As observed by Bermúdez-Otero (2012):

[O]ne should countenance the possibility that variations in lexical experience across speakers may manifest themselves as differences in the extent to which individual learners acquire morphophonological generalizations over stem-level items, including whether or not they end up building parallel cophonologies: **some** speakers may acquire **some** aspects of stem-level phonology late, or not at all.

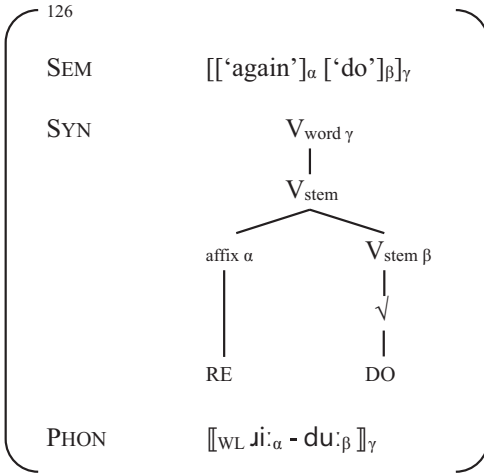
(Bermúdez-Otero 2012: 76; emphasis in the original)

#### 4. The structure of lexical entries

The evidence that we have reviewed can be taken to mean that most language users eventually parse all prefixed words into a prefix and a base, although this may not necessarily occur upon the first encounter with the word and that lexical entries may be reanalysed as a result of comparisons with other lexical entries. Obviously, the result of that parsing would be different depending on the semantic transparency of the prefixed word. If we adopt the framework of Stratal Phonology (Bermúdez-Otero 2018), words that have productive prefixes go through word-level phonology and may be stored analytically, or not at all (and so be produced only ‘on-line’).<sup>18</sup> If stored, the lexical entry of a word such as *re-do* would look something like (12).

<sup>18</sup>This is not meant to exclude other approaches which also allow for those storage of complex words (e.g. Construction Morphology) and may thus also be well-suited to deal with such cases. Note that the binary stratal split in Stratal Phonology is not necessarily incompatible with the kind of gradience in the types of prefixed words that we have been describing. In our view, most of the gradience lies either in the likelihood of segmentation and, if there is segmentation, in the degree of opacity of the formatives involved, and need not be modelled in the phonology itself.

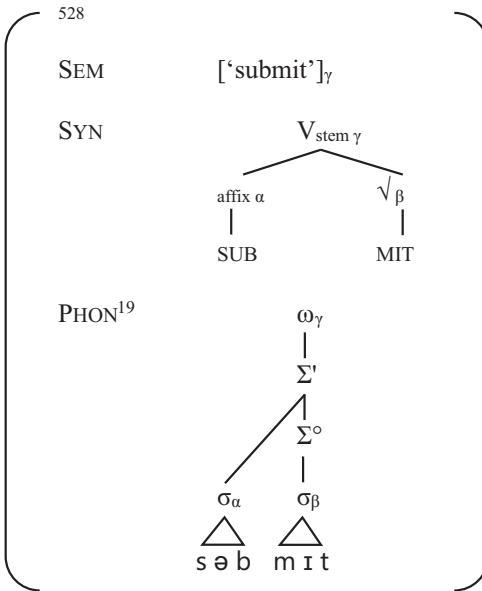
(12) Lexical entry for *re-do*



This kind of representation is borrowed from Bermúdez-Otero (2013) and Jackendoff (1975). It uses two key ideas: morphologically complex words (or phrases) can be analytically listed in the lexicon (Bermúdez-Otero 2012), i.e. in a decomposed form, and different levels of representation within a lexical entry can be coindexed (Jackendoff 1997, 2002). In (12), we represent coindexation using postsubscripted Greek letters ( $\alpha$ ,  $\beta$  and  $\gamma$ ). Here, all levels of representation are coindexed with one another. We voluntarily leave the semantic level relatively vague, as there are quite diverse ways to approach lexical semantics (see Acquaviva et al. (2020) for an overview). Figure 126 is an arbitrarily chosen number representing the index of the entry.

The evidence reviewed in previous sections points to complexity at the morphological level despite the absence of complexity at the semantic level. Therefore, we propose that the lexical entries of opaque prefixed words are stored as in (13), with the example of the word *submit*.

(13) **Lexical entry for *submit***



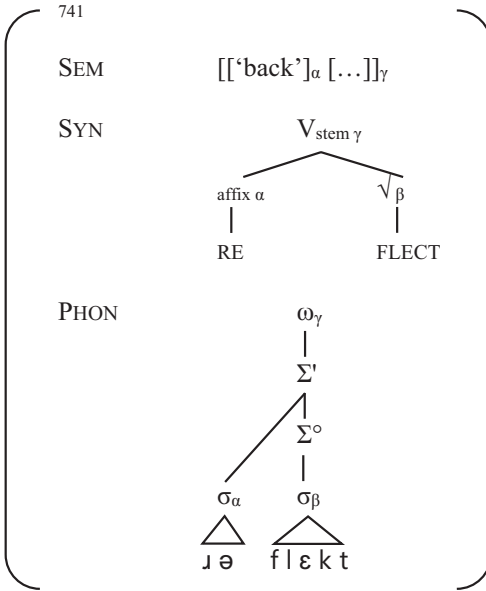
Here, crucially, coindexation relates morphological constituents with their phonological realization ( $\alpha$  and  $\beta$ ), but the semantic representation is not related to those constituents and is directly related to the whole word on other levels ( $\gamma$ ). Note that the structure present at the morphological level may be accessed by the morphology, e.g. to know which nominalizing suffix should be used for a given root. In Stratal Phonology, such an entry is typical of ‘nonanalytic listing’, i.e. ‘with their stem-level phonological properties fully specified’ (Bermúdez-Otero 2012, after Jackendoff 1975). Such representations could explain why words like *submit* behave like words with productive morphology in psycholinguistic experiments that neutralize meaning but not when those experiments do not neutralize meaning (see §2.3.1). Note that nonanalytic listing does not imply that there cannot be internal complexity at the morphological or semantic levels, although an absence of complexity or compositionality may be a reason to store the word nonanalytically in the first place.

As we saw in §2.4 and in the previous section, a prefix or a root may have recurrent meaning that can be quite underspecified. Assuming that language users are able to detect such subtle semantic or thematic regularities like the ones identified by Baeskow (2006), the semantic representation of *submit* might be reanalysed as something as including thematic

<sup>19</sup> We are here using recursive foot structure (with  $\Sigma^{\circ}$  representing a minimal foot projection and  $\Sigma'$  representing a maximal foot projection), in line with Jensen (2000) and Davis & Cho (2003), although it is probably possible to assume a structure like  $(səb(mit)_{\Sigma(\omega^{\circ})})_{\omega'}$  in which the first syllable is excluded from a minimal phonological word projection ( $\omega^{\circ}$ ) and adjoined to a higher projection ( $\omega'$ ) (for an analysis similar to this, see Raffelsiefen (2007)). The debate exists for ‘neutral’ suffixes (see Bermúdez-Otero (2011)) and it can probably only be settled using durational evidence. The central element is that prefixes should not belong to  $\Sigma^{\circ}$ , and so we leave the details of the rest of the prosodic analysis to future research.

information such as that in (6). Likewise, a word in which the prefix contributes meaning may be stored as (14), in which only the prefix is coindexed to part of the semantic representation. Similar representations could be posited for words in which the root contributes meaning to the whole, but not the prefix (see (8)).

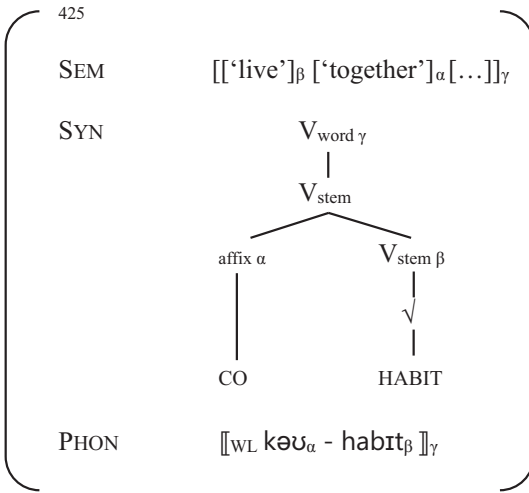
(14) Lexical entry for *reflect*



Cases such as those seen in (9) and (10) in which the prefix contributes meaning but is attached to a bound stem may be analysed as word-level constructions if the semantic contribution of the prefix is perceived to be the same as that found in constructions with free bases, and the base can be attributed a syntactic category. Thus, the lexical entry for a word like *cohabit* could be analysed as in (15).<sup>20</sup> We represent the fact that the meanings of the prefix and the stem are perceived through coindexation, shown as  $\alpha$  and  $\beta$ , but indicate '[...]' at the semantic level to show that there are semantic idiosyncrasies that are specific to that word (e.g. 'living together **as if married**' – first sense given by Dictionary.com).

<sup>20</sup> In English, stems are usually free, but Bermúdez-Otero (personal communication) suggests that items such as *flute* may be analysed as bound stems, and this is the analysis that we are adopting for *habit*.

**(15) Lexical entry for *cohabit***



Once again, analytic listing does not necessarily mean that word-level formations are compositional at all levels of representation, but that words that are analytically listed are computed as constructs at the phonological level. For instance, compounds are treated as word-level constructs, and it is well-established that they may drift away semantically from the meaning of their constituents. Possibly the most extreme case of this is place names in which the semantics of the constituents are underspecified or entirely absent (e.g. *Cambridge, Washington, Westminster*), even though they often have the phonological properties of word-level constructions (see Dabouis 2024, Köhnlein 2015). This is also true of some phrase-level constructions such as phrasal idioms (e.g. *to pull X’s leg*) which have non-compositional semantics and thus require lexical storage (Bermúdez-Otero 2012).

On top of having stored stems and words which have constituent structures involving prefixes and bound roots, language users may posit lexical entries for those constituents, as the lexicon is assumed to be highly redundant in Stratal Phonology. Thus, those entries may specify recurrent semantic features, allomorphs or the preferred nominalizing suffix for roots.

We have already mentioned the idea that the generalizations that are related to opaque prefixed words are lexical redundancy rules that are typical of stem-level phonology. Such generalizations are typically semiproductive and sustain lexical exceptions. To capture most of the phonological generalizations detailed in §2.1, we could use an optimality-theoretic framework (Prince & Smolensky 1993) using a constraint requiring the alignment of a minimal foot projection with the left edge of the root (to account for root-initial primary stress in verbs, adjectives and adverbs and for the absence of reduction – arguably secondary stress – in trochaic disyllabic nouns) and one requiring that prefixes are not footed (to account for vowel reduction and the overall ‘stress-repellent’ behaviour found for secondary stress and in verbs, adjectives and adverbs).<sup>21</sup> We would also require

<sup>21</sup> One of the facts exposed earlier in the paper constitutes a potential challenge for Stratal Phonology. Indeed, the theory assumes that morphosyntactic information is lost after the phonology has applied at the stem-level, and so it

machinery to capture the difference between nouns and other syntactic categories. One solution would be to use cophologies, as proposed by Bermúdez-Otero & McMahon (2006), in which the two constraints that refer to prefixes and roots would have different rankings or weights. A representational solution could also be considered. For example, Hammond (1999: 278–82) assumes that unsuffixed verbs and adjectives end in a catalectic suffix. Only a detailed exploration of those two solutions based on large datasets can help determine which analysis should be favoured. In both cases, the two constraints would have to be dominated by other constraints, for example, to enforce secondary stress on the prefix if two syllables precede primary stress (e.g. *còrrespònd*, *èntertàin*, *ìnterséct*) or if a stem-level suffix requires primary stress to be on the prefix (e.g. *àdvocate*, *démonstrate*, *sùbjugate*). Given the non-categorical distributions of the data discussed in §2.1, probabilistic models using weighted constraints (e.g. Goldwater & Johnson 2003, Pater 2009b) could be the best tools to capture those distributions. We leave the details of such an analysis to future research.

## 5. Conclusion

In the phonological literature, English opaque prefixed words are usually seen as simplex words and their internal structure is seen as unlearnable. We have seen that this view, which would predict similar patterns for such words and words without any morphological structure, is unsupported by the available evidence. These words display phonological behaviours that are distinct from those of non-prefixed words, have specific allomorphies and selectional restrictions, appear to be treated like words with free morphemes in early visual decomposition and impact stress placement in the reading of disyllabic words. Moreover, we have seen that, while many prefixed words have non-compositional semantics, even the most opaque roots may actually have ‘skeletal’ thematic information, as argued by Baeskow (2006).

We have addressed the issue of learnability by arguing that language users probably use the same mechanisms to detect opaque morphological structures as the ones that they use to detect productive morphological structures, the main one being systematic comparisons across classes of words sharing form and, at least sometimes, meaning. However, given the more restricted – especially semantic – overlap across words sharing the prefixes and roots discussed in this paper, we have argued that it is possible that not all language users eventually detect such constituents and that those that do probably do it late.

We have also proposed what the lexical entries of different types of prefixed words could look like, and we have appealed to the notions of analytic and nonanalytic listing from Stratal Phonology and coindexation among different levels of representation in lexical entries. In such an approach, there need not be coindexation between all levels of representation, and so items may be semantically non-compositional but complex at other levels of representation. We have suggested that the generalizations required to capture the phonological behaviour of opaque prefixed words are lexical redundancy rules that are typical of stem-level phonology.

---

would not predict that the initial syllable of stem-level derivatives would behave differently if there was a stress shift. However, as we saw in § 2.1.1, prefixed derivatives have reduced vowels significantly more often than non-prefixed ones.

These hypotheses on synchronic analysis predict two types of mismatches with etymology: historically non-prefixed words treated as prefixed, and historically prefixed words no longer perceived as such (see Fournier (1996)). The boundary between opaque prefixed words and truly simplex words should also be investigated in more detail.

It seems to us that the next steps are to continue accumulating phonological, morphological and psycholinguistic evidence which may inform us on how language users process opaque prefixed words. In particular, more evidence from psycholinguistics would be welcome, as a lot of the evidence on morpho–orthographic segmentation comes from studies on suffixed words, and we have reviewed the only studies that we know of which deal with prefixed words. Therefore, more evidence on prefixed words using various experimental paradigms (lexical decision, ERP, segment shifting) would help consolidate existing findings. Moreover, it will also be crucial for future studies to integrate the diversity of the semantic configurations found in prefixed words because, as we have seen, the semantic contribution of prefixes and roots can vary drastically from one word to another. Promising avenues of research may also be to explore more subtle semantic similarities among words sharing prefixes and roots using Distributional Semantics and semantic vectors (Widdows & Cohen 2010) and, on the formal side, to evaluate the contribution of opaque prefixes and roots to computational models modelling stress, such as analogical models or Naive Discriminative Learning (Baayen et al. 2011). Finally, it would also be interesting to investigate individual differences between language users and to establish which factors seem to affect how likely they are to process opaque prefixed words like complex words (e.g. vocabulary size, morphological sensitivity, age, education).

As for models of the morphology–phonology interface, the evidence that we have reviewed shows that reference to semantically opaque constituents should be allowed. Any model which seeks to maintain the view that opaque prefixed words are simplex words should provide an alternative account for the results reported in this paper. The evidence discussed in this paper shows that the range of possible morphological structures that may impact phonology should be expanded to include less productive and semantically transparent structures, possibly in other languages (at least languages closely related to English such as Dutch or German). There is already an emerging line of research on place names (Dabouis 2024, Köhnlein 2015, Mascaró 2016), which shows clear phonological evidence of morphological complexity (e.g. stress patterns that are typical of complex words, deviant phonotactics) despite the semantic opacity of word-internal constituents. In that case too, the recurrence of the internal constituents of proper names, especially final constituents, probably plays a key role in the identification of internal structure.

**Acknowledgements.** We would like to thank two anonymous reviewers for their constructive remarks and suggestions, which have helped us to considerably improve this paper. We would also like to thank the colleagues with whom we have exchanged extensively about those issues over the past few years, most notably Ricardo Bermúdez-Otero, the members of the ERSaF project, Sabine Arndt-Lappe, Ingo Plag, Marie Gabillet and Aaron Seiler, and the audiences of the 2017 French Phonology Network conference (RFP) and Old World Conference in Phonology (OCP) conferences, to whom early versions of this research have been presented. All errors are ours alone.

**Funding.** This research was supported by funding from the Agence National de la Recherche (ANR-21-FRAL-0001-01) and the Deutsche Forschungsgemeinschaft (AR 676/3-1) for the ERSaF (English Root Stress across Frameworks) project.

**Competing interests.** The authors declare none.



## References

- Acquaviva, Paolo, Alessandro Lenci, Carita Paradis & Ida Raffaelli. 2020. Models of lexical meaning. In Vito Pirelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage*, 353–404. Berlin & Boston: de Gruyter Mouton. doi:10.1515/9783110440577-010.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Arndt-Lappe, Sabine & Dabouis, Quentin. Submitted. Secondary stress, lexical storage, and morphological structure – new evidence from dictionary and speech data.
- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Baayen, Harald R, Petar Milin, Dusica Filipović Durdević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118. 438–482. doi:https://doi.org/10.1037/a0023851.
- Baeskow, Heike. 2002. *Abgeleitete personenbezeichnungen im Deutschen und Englischen. Abgeleitete personenbezeichnungen im Deutschen und Englischen*. Berlin: Walter de Gruyter. doi:10.1515/9783110904581.
- Baeskow, Heike. 2004. *Lexical properties of selected non-native morphemes of English*. Tübingen: Gunter Narr Verlag.
- Baeskow, Heike. 2006. A revival of romance roots. *Morphology* 16(1). 3–36. doi:10.1007/s11525-006-0001-7.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English Morphology*. Oxford: Oxford University Press.
- Becker, Michael, Nihan Ketez & Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125. doi:10.1353/lan.2011.0016.
- Bermúdez-Otero, Ricardo. 2011. Cyclicity. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology (vol. 4: Phonological interfaces)*, 2019–2048. Malden, MA: Wiley-Blackwell.
- Bermúdez-Otero, Ricardo. 2012. The architecture of grammar and the division of labour in exponence. In Jochen Trommer (ed.), *The phonology and morphology of exponence - the state of the art*, 8–83. Oxford: OUP.
- Bermúdez-Otero, Ricardo. 2013. The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. *Probus* 25(1). 3–103. doi:https://doi.org/10.1515/probus-2013-0009.
- Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford: Oxford University Press.
- Bermúdez-Otero, Ricardo & April McMahon. 2006. English phonology and morphology. In Bas Aarts & April M S McMahon (eds.), *The handbook of English linguistics*, 382–410. Oxford: Blackwell.
- Beyersmann, Elisabeth & Jonathan Grainger. 2023. The role of embedded words and morphemes in reading. In David Crepaldi (ed.), *Linguistic morphology in the mind and brain*, 26–49. Abingdon & New York: Routledge. doi:10.4324/9781003159759-3.
- Booij, Geert E. 2016. Construction morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 424–448. Cambridge: Cambridge University Press.
- Burzio, Luigi. 1994. *Principles of English stress*. New York: Cambridge University Press.
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Clark, Eve V. 2014. Acquisition of derivational morphology. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 424–439. Oxford: Oxford University Press.
- Coch, Donna, Jennifer Bares & Allison Landers. 2013. ERPs and morphological processing: The N400 and semantic composition. *Cognitive, Affective and Behavioral Neuroscience* 13(2). 355–370. doi:10.3758/s13415-012-0145-3.
- Coch, Donna, Jianjun Hua & Allison Landers-Nelson. 2020. All morphemes are not the same: accuracy and response times in a lexical decision task differentiate types of morphemes. *Journal of Research in Reading*, 43(3). 329–346. doi:10.1111/1467-9817.12306. https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9817.12306 (10 October, 2022).
- Collie, Sarah. 2007. *English stress preservation and stratal optimality theory*. Edinburgh: University of Edinburgh dissertation.
- Creemers, Ava. 2023. Morphological processing in spoken-word recognition. In David Crepaldi (ed.), *Linguistic morphology in the mind and brain*, 50–64. Abingdon & New York: Routledge. doi:10.4324/9781003159759-4.
- Dabouis, Quentin. 2018. The pronunciation of vowels with secondary stress. *Corela*, 16(2). doi:https://doi.org/10.4000/corela.7153.

- Dabouis, Quentin. 2020. Secondary stress in contemporary British English: An overview. *Anglophonia* (30). doi: <https://doi.org/10.4000/anglophonia.3476>.
- Dabouis, Quentin. 2024. Phonological evidence for morphological complexity in English proper names. *Anglophonia* (36). <https://journals.openedition.org/anglophonia/5390>.
- Dabouis, Quentin & Jean-Michel Fournier. Submitted. An empirical study of vowel reduction and preservation in British English.
- Dabouis, Quentin & Jean-Michel Fournier. 2023. The stress patterns of English verbs: Syllable weight or morphology? In Nicolas Ballier, Jean-Michel Fournier, Anne Przewozny & Eiji Yamada (eds.), *New perspectives on English word stress*, 154–191. Edinburgh: Edinburgh University Press.
- Dahak, Anissa. 2011. *Etude diachronique, phonologique et morphologique des syllabes inaccentuées en anglais contemporain*. Paris: Université de Paris Diderot dissertation.
- Davis, Stuart & Mi-Hui Cho. 2003. The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics* 41(4). 607–652. doi: <https://doi.org/10.1515/ling.2003.020>.
- Dawson, Nicola, Kathleen Rastle & Jessie Ricketts. 2018. Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning Memory and Cognition* 44(4). 645–654. doi: [10.1037/xlm0000485](https://doi.org/10.1037/xlm0000485).
- Diependaele, Kevin, Dominiek Sandra & Jonathan Grainger. 2009. Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition* 6(37). 895–908. doi: <https://doi.org/10.3758/MC.37.6.895>.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619. doi: [10.2307/415037](https://doi.org/10.2307/415037).
- Emmorey, Karen D. 1989. Auditory morphological priming in the lexicon. *Language and Cognitive Processes* 4(2). 73–92. doi: <https://doi.org/10.1080/01690968908406358>.
- Fejzo, Anila, Alain Desrochers & S. Hélène Deacon. 2018. The acquisition of derivational morphology in children. In Rachel Berthiaume, Daniel Daigle & Alain Desrochers (eds.), *Morphological processing and literacy development*, 112–132. New York: Routledge. doi: [10.4324/9781315229140-5](https://doi.org/10.4324/9781315229140-5).
- Feldman, Laurie Beth, Patrick A O'Connor & Fermín M Del Prado Martín. 2009. Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review* 16(4). 684–691. doi: [10.3758/PBR.16.4.684](https://doi.org/10.3758/PBR.16.4.684).
- Fidelholtz, J. 1975. Word frequency and vowel reduction in English. *Chicago Linguistic Society* 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(4–5). 539–561. doi: [10.1080/01690960050119698](https://doi.org/10.1080/01690960050119698).
- Fournier, Jean-Michel. 1996. La reconnaissance morphologique. *8ème Colloque d'Avril sur l'anglais oral*, 45–75. Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV.
- Fudge, Erik. 1984. *English word stress*. London: G. Allen & Unwin.
- Gaffiot, Félix. 1934. *Dictionnaire illustré Latin-Français*. Paris: Hachette.
- Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spender, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm workshop on variation within optimality theory*, 111–120. Stockholm: Stockholm University.
- Guierre, Lionel. 1979. *Essai sur l'accentuation en anglais contemporain : Eléments pour une synthèse*. Paris: Université Paris-VII dissertation.
- Guierre, Lionel. 1990. Mots composés anglais et agrégats consonantiques. *5ème Colloque d'Avril sur l'anglais oral*, 59–72. Villetaneuse: Université de Paris-Nord, CELDA, diffusion APLV.
- Halle, Morris & Michael Kenstowicz. 1991. The free element condition and cyclic versus noncyclic stress. *Linguistic Inquiry* 22(3). 457–501.
- 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.
- Hammond, Michael. 1989. Cyclic secondary stresses in English. *Proceedings of the annual meeting of the West Coast Conference on Formal Linguistics*, vol. 8, 139–153. Stanford, California: Stanford Linguistics Association, Stanford University.
- Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach*. (Ed.) Jacques Durand. Oxford: Oxford University Press.
- Hammond, Michael. 2003. Frequency, cyclicity, and optimality. University of Arizona. <http://www.u.arizona.edu/~hammond/kslides.pdf> (accessed 2 September, 2014).

- Harley, Heidi. 2009. A morphosyntactic account of the “Latinate” ban on dative shift in English. Paper presented at the colloquium of the Linguistics Department at the University of California, Santa Cruz (15 May).
- Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3/4). 225–276. doi:[10.1515/tl-2014-0010](https://doi.org/10.1515/tl-2014-0010).
- Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13(2). 227–276.
- Heyer, Vera & Dana Kornishova. 2018. Semantic transparency affects morphological priming ... eventually. *The Quarterly Journal of Experimental Psychology* 71(5). 1112–1124. doi:<https://doi.org/10.1080/17470218.2017.1310915>.
- Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671. doi:<https://doi.org/10.2307/412891>.
- Jackendoff, Ray. 1997. *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jensen, John T. 2000. Against ambisyllabicity. *Phonology* 17(2). 187–235. doi:[10.1017/S0952675700003912](https://doi.org/10.1017/S0952675700003912).
- Katamba, Francis & John T. Stonham. 2006. *Morphology. Modern linguistics series*. 2nd ed. Houndmills, Basingstoke, Hampshire; New York, NY: Palgrave Macmillan.
- Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10(3). 421–441.
- Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In Harry Van der Hulst & Norval Smith (eds.), *The structure of phonological representations I*, 131–175. Dordrecht: Foris. doi:<https://doi.org/10.1515/9783112328088-008>.
- Ktori, Maria, Petroula Mousikou & Kathleen Rastle. 2018. Cues to stress assignment in reading aloud. *Journal of Experimental Psychology* 147(1). 36–61. doi:[10.1037/xge0000380](https://doi.org/10.1037/xge0000380).
- Ktori, Maria, Jeremy J. Tree, Petroula Mousikou, Max Coltheart & Kathleen Rastle. 2016. Prefixes repel stress in reading aloud: Evidence from surface dyslexia. *Cortex* 74 (October 2015). 191–205. doi:[10.1016/j.cortex.2015.10.009](https://doi.org/10.1016/j.cortex.2015.10.009).
- Laudanna, Alessandro, Cristina Burani & Antonella Cermele. 1994. Prefixes as processing units. *Language and Cognitive Processes* 9(3). 295–316. doi:[10.1080/01690969408402121](https://doi.org/10.1080/01690969408402121).
- Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(01). 1–75. doi:[10.1017/S0140525X99001776](https://doi.org/10.1017/S0140525X99001776).
- Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.
- Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.
- Marelli, Marco, Daniela Traficante & Cristina Burani. 2020. Reading morphologically complex words: Experimental evidence and learning models. In Vito Pirelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*. 553–592. Berlin & Boston: de Gruyter Mouton. doi:[10.1515/9783110440577-014](https://doi.org/10.1515/9783110440577-014).
- Marslen-Wilson, William D., Mirjana Bozic & Billi Randall. 2008. Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes* 23(3). 394–421.
- Marslen-Wilson, William D., Lorraine Komisarjevsky Tyler, Rachele Waksler & Lianne Older. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review* 101(1). 3–33. doi:[10.1037/0033-295X.101.1.3](https://doi.org/10.1037/0033-295X.101.1.3).
- Mascaró, Joan. 2016. Morphological exceptions to vowel reduction in central catalan and the problem of the missing base. *Catalan Journal of Linguistics* 15. 27–51. doi:[10.5565/rev/catjl.184](https://doi.org/10.5565/rev/catjl.184).
- McKinnon, Richard, Mark Allen & Lee Osterhout. 2003. Morphological decomposition involving non-productive morphemes: ERP evidence. *NeuroReport* 14(6). 883–886. doi:<https://doi.org/10.1097/00001756-200305060-00022>.
- Melinger, Alissa. 2001. *Morphological complexity in English prefixed words: an experimental investigation. Methodology*. Buffalo, NY: University of New York at Buffalo dissertation.
- Minkova, Donka & Z.L. Zhou. 2022. Early metrical and lexicographical evidence for functional stress-shifts. *English Language & Linguistics*. Cambridge University Press 26(3). 533–558. doi:[10.1017/S1360674322000144](https://doi.org/10.1017/S1360674322000144).
- Moore-Cantwell, Claire. 2016. *The representation of probabilistic phonological patterns: neurological, behavioral, and computational evidence from the English stress system*. Amherst, MA: University of Massachusetts Amherst dissertation.
- Mousikou, Petroula, Jasmin Sadat, Rebecca Lucas & Kathleen Rastle. 2017. Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. *Journal of Memory and Language*. Academic Press 93. 169–192. doi:[10.1016/J.JML.2016.09.003](https://doi.org/10.1016/J.JML.2016.09.003).

- Mudgan, Joachim. 2015. Units of word-formation. In P.O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe. Volume 3*, 235–301. Berlin, Boston: Mouton de Gruyter.
- Nessly, Larry. 1982. Vowel tenseness in English: Another look. *Language Learning* 32(2). 393–410.
- Pastizzo, Matthew J. & Laurie B. Feldman. 2004. Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language* 90(1–3). 31–39. doi:10.1016/S0093-934X(03)00417-6.
- Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17(2). 237–274. doi:https://doi.org/10.1017/S0952675700003900.
- Pater, Joe. 2009a. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. *Advances in Optimality Theory: Phonological Argumentation. Essays on Evidence and Motivation*, 123–154. London: Equinox. doi:https://doi.org/10.7282/T3FX77CX.
- Pater, Joe. 2009b. Weighted constraints in generative linguistics. *Cognitive Science* 33(6). 999–1035. doi:https://doi.org/10.1111/j.1551-6709.2009.01047.x.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Poldauf, Ivan. 1984. *English word stress*. Oxford: Pergamon Press.
- Prince, Alan & Paul Smolensky. 1993. *Optimality theory. Constraint interaction in generative grammar*. New Brunswick; Bolder: Rutgers University.
- Raffelsiefen, Renate. 1993. Relating words: A model of base recognition. Part I. *Linguistic Analysis* (23). 3–161.
- Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. *Phonology* 21(1). 91–142. doi:10.1017/S0952675704000144.
- Raffelsiefen, Renate. 2007. Morphological word structure in English and Swedish: The evidence from prosody. *Fifth Mediterranean Morphology Meeting*, 209–268.
- Raffelsiefen, Renate. 2015. Phonological restrictions on English word-formation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, 894–917. Berlin, Boston: Mouton de Gruyter.
- Rastle, Kathleen. 2019. The place of morphology in learning to read in English. *Cortex* 116. 45–54. doi:10.1016/j.cortex.2018.02.008.
- 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. doi:https://doi.org/10.1006/jmla.1999.2687.
- Rastle, Kathleen & Matthew H Davis. 2008. Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes* 23(7–8). 942–971.
- Ravid, Dorit, Emmanuel Keuleers & Wolfgang U. Dressler. 2020. Emergence and early development of lexicon and morphology. In Vito Pirelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*, 593–633. Berlin & Boston: de Gruyter Mouton. doi:10.1515/9783110440577-015.
- Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11(3). 563–605.
- Siegel, Dorothy C. 1974. *Topics in English Morphology*. Cambridge, MA: MIT dissertation.
- Siegel, Dorothy C. 1980. Why there is no = boundary. In Mark Aronoff & Mary-Louise Kean (eds.), *Juncture: A Collection of Original Papers*, 131–134. Saratoga: Anma Libri.
- Sonderegger, Morgan & Partha Niyogi. 2013. Variation and change in English noun/verb pair stress: Data and dynamical systems models. In A.C.L. Yu (ed.), *Origins of Sound Change*, 262–284. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199573745.003.0013.
- Stanners, Robert F., James J. Neiser & Scott Painton. 1979. Memory representation for prefixed words. *Journal of Verbal Learning and Verbal Behavior* 18(6). 733–743. doi:10.1016/S0022-5371(79)90439-0.
- Taft, Marcus. 1979. Recognition of affixed words and the word frequency effect. *Memory & Cognition* 7(4). 263–272. doi:10.3758/BF03197599.
- Taft, Marcus. 1994. Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes* 9(3). 271–294. doi:10.1080/01690969408402120.
- Taft, Marcus. 2003. Morphological representation as a correlation between form and meaning. In Egbert M. H. Assink & Dominiek Sandra (eds.), *Reading complex words*, 113–137. New York: Springer Science+Business Media. doi:10.1007/978-1-4757-3720-2\_6.
- Taft, Marcus & Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14(6). 638–647. doi:10.1016/S0022-5371(75)80051-X.
- 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.

- Taft, Marcus & Paul Kougiouss. 2004. The processing of morpheme-like units in monomorphemic words. *Brain and Language* 90(1–3). 9–16. doi:10.1016/S0093-934X(03)00415-2.
- Treiman, Rebecca, Nicole Rosales, Lauren Cusner & Brett Kessler. 2020. Cues to stress in English spelling. *Journal of Memory and Language* 112. doi:10.1016/j.jml.2020.104089.
- Tyler, Andrea & William Nagy. 1989. The acquisition of English derivational morphology. *Journal of Memory and Language* 28(6). 649–667. doi:10.1016/0749-596X(89)90002-8.
- Wells, John C. 2008. *Longman Pronunciation Dictionary*. 3rd ed. London: Longman.
- Widdows, Dominic & Trevor Cohen. 2010. The semantic vectors package: New algorithms and public tools for distributional semantics. *Proceedings - 2010 IEEE 4th International Conference on Semantic Computing, ICSC 2010*. 9–15. doi:10.1109/ICSC.2010.94.
- Wurm, Lee H. 1997. Auditory processing of prefixed English words is both continuous and decompositional. *Journal of Memory and Language* 37(3). 438–461. doi:10.1006/jmla.1997.2524.
- Wurm, Lee H. 2000. Auditory processing of polymorphemic pseudowords. *Journal of Memory and Language* 42. 255–271. doi:10.1006/jmla.1999.2678. <https://doi.org/10.1006/jmla.1999.2678>.
- Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28(2). 417–472. doi:10.1007/s11049-010-9095-z.