
The phonotactics of some English morphology

Laurie Bauer
Victoria University of Wellington, New Zealand

1. Introduction

I recall vividly, from my years working in Odense, having conversations with Hans Basbøll on the subjects both of phonological gemination and of morphological productivity. It is thus particularly satisfying to be able to offer a brief study for this Festschrift in which these two topics are brought together.

In monomorphemic words of English, there are strict restrictions on the cooccurrence of identical segments. We can phrase some of these as constraints in the tradition of Optimality Theory as in (1).

\[\begin{align*}
(a) \quad & \ast \text{ continuant} \quad \text{nasal} \\
(b) \quad & \ast \text{ sonorant} \quad \text{continuant} \\
(c) \quad & \ast \text{ consonant} \quad \text{sonorant} \\
(d) \quad & \ast \text{ vocalic} \quad \text{sonorant} \\
\end{align*}\]
Of these, (1e) requires some comment. First, I am assuming an analysis in which diphthongs have two vocalic nuclei under the same V node, so that the vowels in boy, cow and sew are not counter-evidence to (1e). Second, (1e) may be broken in some non-native words such as Malvolio (/ˈmævləʊ/) and onomatopoeia (/ɒnəmatəˈpɔiə/). In some varieties of English, (1e) is broken rather more regularly with words like cure pronounced /kjuːə/ (in two syllables; similarly for words in the near and square lexical sets), but this is not true for RP under standard analyses.

A number of inflectional morphemes of English have several allomorphs in order to maintain some of the constraints listed in (1). For example, the /ɪd/ allomorph of the past tense or past participle allows (1a) to be maintained for verb-stems ending in /d/ (e.g. melded), while the /ɪz/ allomorph of the plural morpheme allows (1b) to maintained for noun stems ending in a /z/ (e.g. adzes). In both cases, the allomorph is used to exclude a wider range of segmental cooccurrences than is excluded by (1), but this paper focuses on the constraints outlined in (1).

Not all of the constraints in (1) are maintained in polymorphemic words. For example freer, truest break (1e), and (1e) is also often broken by the addition of -ing, for example in words like being, seeing and going.

This paper considers how the constraints in (1) fare in English derivational morphology. In section 2 some phonological preliminaries are discussed, then sections 3 and 4 deal with prefixes and suffixes respectively, before some tentative conclusions are drawn in section 5.
2. Some phonological preliminaries

2.1 The happy vowel

The phonological status of the happy vowel is much discussed in English. In some varieties, speakers associate it with the same phoneme that they have in the KIT lexical set; in others it is associated with the same vowel phoneme as occurs in the FLEECE lexical set; in some of these varieties its phonetic realisation may be distinct from the typical realisations of the phoneme it is most often associated with, and in some varieties its realisation may be as open as that of the DRESS vowel (Stoddart, Upton & Widdowson 1999:75). The assumption is made here that it should be seen as phonologically distinct from the vowels in both KIT and FLEECE, and it is transcribed /i/ (i.e. as a short vowel). It will be shown in section 5 that this assumption can be justified in the light of this research.

2.2 Non-pre-vocalic /r/

This paper deals with the situation found in non-rhotic varieties of English. This means that words like air and prefixes like per- do not end in an /r/, and there is no point in asking whether they can be followed by morphological elements beginning in /r/. At the same time, there is no point in asking whether they can be followed by vowel-initial elements, because the moment they are, a linking /r/ appears, and there is no vowel sequence. Thus perorate cannot have two abutting vowels. This cuts down the number of potential violations of the constraints in (1).

2.3 Other potential links

In a similar vein, it could be argued that vowels such as /əu/ and /au/ produce a linking /w/, and vowels such as /əi/ and /ɔi/ produce a linking /j/ which has a similar effect when morphological elements ending in those segments
occur before vowel-initial elements. This is not the position taken here, with the result that co-ownership is analysed as a violation of (1d) over the morpheme boundary. If co-ownership were analysed as not violating (1d) (because it contained the sequence /əuəu/), then (1d) and (1e) would be almost completely vacuous over boundaries, because nearly all potential input vowels would be followed by a linking consonant when the next morphological element began with a vowel.

3. Prefixes

The list of prefixes from Marchand (1969) was taken as the fundamental data base for this study, and the bases they cooccur with were discovered from Allen (2000). Both of these procedures lead to suggestive rather than definitive answers. In particular, Marchand lists very few “prefixes” ending in /əu/ (like neo-). These combining forms are likely to be prefixed to a wider range of roots when the entire lexicon is considered.

Some prefixes have prevocalic allomorphs: contrast, for instance, athematic with anechoic. Combining forms, too, lose the final -o before vowel-initial combining forms so that we find pseudo-carp but pseudonym not *pseudo-name (for some discussion see Bauer and Huddleston 2002:1662ff). However, in more modern formations, the -o is often retained before vowel-initial stems (e.g. pseudo-intellectual).

Vowel-final prefixes can end in one of five vowels: /ə/, /i/, /i:/, /əu/ and /ai/. There is only one instance of (1d) being violated in the data (co-ownership, already cited), but the patterns where (1e) was violated were not random. The occurrent patterns are shown in Table 1. It can be seen that short vowels (except /u/ which never occurs initially) occur readily following a prefix and that long vowels and diphthongs, while not excluded, are rare. This is probably due to a shortage of vowel-initial bases with long vowels or
diphthongs: a brief sample from the letter A in Allen (2000) suggests that the ratio of short vowels to long vowels and diphthongs in base-initial position is something like 7:1.

** Table 1 about here **

With consonant-final prefixes, all of (1a, b, c) were violated, as is shown in the examples post-tonic, misspell and unnamed, all pronounced with geminates. Although there were prefixes (such as pan- and mal-) which did not occur in the data with geminate consonants over their boundaries, this has to be seen as a comment on the rarity of the prefixes rather than as a strong constraint.

However, geminates are not necessarily the rule. Whereas Jones (1977) and Wells (1990) list misspell as having a geminate over the boundary, dissatisfaction may have a geminate or a single /s/ and dissimulation is not listed with a geminate at all. The same applies to unnatural, innavigable and innumerable respectively. It seems that geminates are maintained with the most productive morphology, but tend to vanish as the morphology becomes less productive and also as a function of high frequency of the derivative.

4. Suffixes

Again a list of suffixes was taken from Marchand (1969), and again that is a bit problematic given that Marchand lists as separate suffixes things which might be considered allomorphs of the same morpheme. However, using Marchand’s list guarantees that the suffixes have not been chosen to fit my particular agenda. Combinations of base and suffix were traced using Lehnert (1971).

Not only are there general allomorphy-causing processes which eliminate some of the potential violators of the rules in (1), there is quite a lot of variation in pronunciation. For instance, valuable may be /vælˈʃəbl/ or /valjuːˈʃəbl/ or
something in between. Continuous may or may not undergo levelling to /ʊə/ over the morpheme boundary.

Attested patterns of base-final vowel and suffix-initial vowel are shown in Table 2. Note that /ɔː/ is allowed as a stem-final vowel because within a word there is not inevitably an intrusive /r/ following the base in words like gnawing, rawish. The number of suffix-initial vowels is relatively small, and not all of them are attested following a vowel. Some of the gaps in Table 2 are accidental, due to the rarity of some relevant suffixes, and indeed, most of them could be interpreted as being accidental. Note, however, that (1d) is not broken in suffixation.

** Table 2 about here **

Where consonants are concerned, geminates are avoided over morpheme boundaries for most suffixes. In some cases, such as with -hood or -ward it would be phonologically impossible to get the consonant base-finally in any case. In cases like -kin, -le, -most, -ship and -ster it would be phonologically possible to get geminates, but these are not attested. Nevertheless, geminates do arise over morpheme boundaries. There are very few attested examples which violate (1a) or (1b) (lord-dom, grief-ful, shelf-ful) but many that break (1c), especially with geminate /l, m, n/, as the following shows.

Geminate /ll/ is regularly found with -less, -like and adverbial -ly following a stressed syllable: goalless, vowel-like, drolly. Single /l/ is found with adjectival -ly and adverbial -ly following an unstressed syllable: rascally, abysmally. Note there are no appropriate examples to clarify what happens with adjectival -ly after a stressed syllable, so the two could act in the same way.

Geminate /mm/ is found with -ment: embalmment.

Geminate /nn/ is found regularly with -ness: keenness, soddenness.

Note that the conditions for a geminate /r/ do not occur in non-rhotic varieties, but that there are in any case no attested potential sites.

Because there is variation in pronunciation, not all the possible geminates are maintained: wholly may or may not be homophonous with holy (see Bauer
2001:82 for data on -ly). As is shown in Table 3, the maintenance of a geminate is linked with low frequency. Although a few words appear to be in the wrong category given their frequency, on the whole, the more frequent a word is, the less likely it is to have a geminate in it.

** Table 3 about here **

5. Discussion and conclusions

Violations of the constraints in (1) appear to signal morphological boundaries — occasionally former morphological boundaries. Words such as Malvolio with a violation of (1e) may thus be perceived as though they have a morphological boundary in them, despite their lack of analysability. Interestingly, the signals do not work the same way for prefixes and suffixes, prefixes allowing far more breaches of the constraints in (1), including (1d), which is not breached by suffixes. Since compounds and syntactic constructions do not appear to be affected by the constraints in (1), this makes prefixes appear to be more compound-like than suffixes, a fact which is of interest given the apparently compound-like lexical behaviour of prefixes (Bauer 2002).

Where violations of the constraints in (1) would otherwise occur, there are a number of options for avoiding them. Forms such as dramatic, Platonic, malarial, uglify illustrate insertion and deletion strategies for avoiding violations of the constraints in (1), even though these are not generalised in predictable ways. Further, where violations of the constraints in (1) do occur (and perhaps particularly violations of constraints (1a-c)), many of the potential geminates seem to be simplified if the relevant words are frequent enough. We can see this as part of the process of lexicalisation. As words become lexicalised, so it becomes unnecessary for speakers to be aware of their internal structure, and so markers of morphological boundaries within the word can be dispensed with. Given this scenario, it is perhaps the retention of geminate /nn/ in words with
the suffix -ness which requires some explanation, rather than the gradual loss of
geminate /ll/ in -ly adverbs. Although productivity might be part of the answer
if these two affixes behaved in the same way, the fact that two of the most
productive suffixes of English behave in different ways is slightly disconcerting.
Perhaps this is in some way related to the nature of the segments involved:
although this is no more than speculation, it might be that it is easier to have
geminates when there is contact between the articulators than when there is not;
alternatively, it might be due to the difference in secondary articulation of
syllable-final and syllable-initial /l/, which makes gemination of this particular
consonant more difficult.

Since suffixation appears not to violate (1d), but the happy vowel is
followed both by /I/ and by /i:/ we have evidence for saying that the happy
vowel is phonologically distinct from these two other phonemes. This was no
more than a working assumption until now.

A simple analysis of what happens at morphological boundaries can thus
tell us things about the segmental inventory of the language we are considering
as well as something about the way the morphology works. It is also noteworthy
that different varieties of English, by giving us subtly different boundary signals,
may be leaving the way open for different interpretations of what was originally
the same structure. This is something which deserves much deeper
consideration.

* This last suggestion is due to Paul Warren, who read and commented on an earlier draft of this
paper, as also did Winifred Bauer. I should like to express my thanks to both of them for their help.


Table 1: Attested cooccurrence patterns of prefix-final vowel and base-initial vowel

<table>
<thead>
<tr>
<th>Prefix-final</th>
<th>Base-initial vowels found</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ø  i  e  æ  å  a  i:  o:  ɔ  y  ø  æ  ø  ø  ø  ø  ø  =</td>
</tr>
<tr>
<td>ø</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
</tr>
<tr>
<td>i:</td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td></td>
</tr>
<tr>
<td>æø</td>
<td></td>
</tr>
</tbody>
</table>


Table 2: Attested cooccurrence patterns of base-final vowel and suffix-initial vowel

<table>
<thead>
<tr>
<th>Base-final</th>
<th>ə</th>
<th>i</th>
<th>e</th>
<th>i</th>
<th>i:</th>
<th>eː</th>
<th>a</th>
<th>ø</th>
<th>eː</th>
</tr>
</thead>
<tbody>
<tr>
<td>ə</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ø</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eː</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aː</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>øː</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>œː</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>œː</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table 3: Geminate maintenance and frequency: adverbial *-ly* following a stressed syllable ending in /l/.

<table>
<thead>
<tr>
<th>Category</th>
<th>Word</th>
<th>Frequency in the British National Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>No information in Jones (1977)</td>
<td>banally</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>smally</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>swelly</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>tally &lt; tall</td>
<td>0</td>
</tr>
<tr>
<td>Geminate maintained according to Jones (1977)</td>
<td>drolly</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>frailly</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>palely</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>solely</td>
<td>1639</td>
</tr>
<tr>
<td></td>
<td>vilely</td>
<td>10</td>
</tr>
<tr>
<td>Variable geminate according to Jones (1977)</td>
<td>dully</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>foully</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>shrilly</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>wholly</td>
<td>2295</td>
</tr>
<tr>
<td>No geminate according to Jones (1977)</td>
<td>fully</td>
<td>8815</td>
</tr>
<tr>
<td></td>
<td>ideally</td>
<td>1175</td>
</tr>
<tr>
<td></td>
<td>really</td>
<td>46470</td>
</tr>
</tbody>
</table>