The phonetics of '*un*'

Jen Hay

This paper analyses words prefixed with *un-* in a large corpus of New Zealand English, in order to test the hypothesis that the difference between retrieved words and productively coined words may be discerned phonetically. The results reveal that factors which tend to facilitate decomposability (boundary-like phonotactics; high frequency of the derived form relative to the base) are associated with phonetically longer prefixes. That is, more decomposable words (e.g. *unburstable*) tend to have longer prefixes than less decomposable words (e.g. *unfortunate*).

In addition, the study reveals an intriguing change over time, in which contemporary New Zealanders seem to be using *un-* less productively than their counterparts born in the 19th and early 20th Century. Contemporary New Zealanders tend to use *un-* slightly less frequently than older New Zealanders, and are more likely to use sequences of *not*+adjective.

1.   Introduction

Most data for research on morphological productivity/creativity (and indeed, morphology in general), has come from intuition, elicitation or experimentation. Recent work has also examined patterns in large corpora, which contain either written texts, or transcriptions of speech events. However, very little work has examined 'spoken morphology', investigating the way in which morphological structure is actually reflected in the phonetics of speech.

This paper is an exploratory study of morphological creativity in a large corpus of spoken New Zealand English, giving special attention to the phonetic realisation of the prefix *un*.[1] This is a large corpus, containing interviews with hun-

---

dreds of New Zealanders born between 1850 and 1971 (see Gordon et al. 2005). The earliest recordings were made in the 1940s. I demonstrate that the length of the prefix is related to the degree to which it is being used as an independent morpheme and that the use of the prefix *un*- has varied over the history of NZE.

The finding that individual words can contain some phonetic evidence regarding the degree to which they are morphologically decomposable has great potential, not only for furthering our understanding of the contexts in which affixed words tend to be creatively coined, but also for the possibility of gaining insight into the extent of productive word-formation in the speech of specific individuals on specific occasions.

2.   Background

It may seem surprising for a book on lexical creativity to include a paper examining details of the phonetics of a single, relatively common affix. What, after all, is creative about the use of a straightforward prefix like un-? Indeed – sometimes nothing. When a speaker uses words like *unfortunate* or *unnecessary* the chances are that they are being retrieved whole from memory, rather than creatively constructed. The listener, too, is likely to understand the meaning of *unfortunate* without necessarily decomposing the word into its constituent elements or analysing it semantically as *un+fortunate*. In some other words, however, the prefix *un*- may carry an important and distinct part of the meaning: rather than being retrieved whole from memory, the word may be created compositionally, and the listener will have to decompose it in order to acquire its meaning. For example, there is likely to be a greater degree of lexical creativity in producing the word *unburstable* than the word *unfortunate*.

While the focus in this paper is on the affix *un*, this study is intended to stand as a test-case for a more general question. Can the difference between retrieved words (e.g. *unfortunate*) and creatively coined words (such as *unburstable*) be discerned phonetically? If so, this would open up a range of research possibilities, enabling us, for example, to closely study the circumstances under which words tend to be retrieved whole or creatively coined.

The particular focus in this paper is on the length of the affix *un*. Examples (1)–(6) are utterances produced by speakers in the Origins of New Zealand Eng-

lish Corpora (described further below). Following each utterance the length of the prefix *un* is given in milliseconds. The utterances are presented in order of increasing length of the prefix. Since these utterances come from different speakers who no doubt have different speaking rates, comparing these raw length measurements may not tell us much at all. However, the variation here is quite large – the prefix in (6) is more than 3.5 times as long as the prefix in (1). This variation suggests that it is worth investigating whether the length of the affix correlates in some sense with the degree to which it is functioning as an independent constituent.

(1)   but I unfortunately . couldn't open the door to look in (.06ms, nasal only – no vowel produced)

(2)   that was very unsatisfactory (.09m)

(3)   and he let me know it in no uncertain terms (.11ms)

(4)   to get an unboring job (.17ms)

(5)   and had many a trying rough trip on the then rough unmetalled tracks (.22ms)

(6)   when you're looking at it from an unbiased point of view (.22ms)

There is some evidence that phonemes which are prone to deletion are less likely to be deleted if they are affixes. For example, studies on final t-deletion in English show that /t/ is more likely to be deleted in monomorphemic words (such as *past*), than when the /t/ is itself an affix (e.g. *jumped*) (Guy 1980, 1991; Labov 1989).

  If an affix comprises a meaningful constituent of a particular word, it is advantageous and perhaps even necessary for that affix to be phonetically salient rather than reduced or deleted. Indeed, in (1), above, the vowel of the affix was not produced at all; the affix was represented by a syllabic nasal. However, one would be surprised to find the same level of reduction in a word such as *unboring* (in 4). In fact, if a listener simply heard '*boring*' preceded by a syllabic nasal, it is not clear that the intended meaning would be available at all. The same line of reasoning leads to the hypothesis that highly compositional, highly creative words may contain phonetically longer prefixes than less compositional, highly lexicalised words.

3.   The storage and processing of affixed words

Before proceeding to investigate whether the degree of reduction of *un* varies in any systematic way, it is first necessary to outline some assumptions about the processing of affixed words. In the previous section I invoked a contrast between

un words which are lexicalised, and those which are creatively coined. However, it is more productive to conceptualise morphological decompositionality as a continuum. At one extreme are simplex words, or words that have undergone semantic drift to the extent that they are no longer seen as being composed of separate morphemes, and so, for all intents and purposes, they are processed as morphologically simple. At the other end of the continuum are words which may never have been encountered before (such as *unburstable*), and are therefore created and perceived as compositional. Most affixed words, however, are likely to fall somewhere between these two extremes. Hay and Baayen (2005) outline a wide range of literature which suggests that morphological complexity is gradient. Affixed words seem to be 'affixed' to varying degrees. Those which have been previously encountered are stored in the lexicon – including words with inflectional affixes. Both retrieval and composition play a role (to varying degrees) in lexical access. A useful metaphor for considering the balance of retrieval and composition in lexical access is the dual route model suggested by Baayen and Schreuder (see, e.g. Baayen 1992; Schreuder and Baayen 1995; Baayen and Schreuder 1999). According to this model, lexical access in perception proceeds as a race between direct access of the 'full form' (e.g. *unhappy*), and parsing of the constituents (*un+happy*). While either route may be solely responsible for access on any given occasion, in general the two routes converge interactively on the correct meaning representation. The relative contribution of the different routes is not simply a matter of lexical access on that particular occasion, but rather the access route impacts the storage of the word. Words for which the compositional route dominates tend to be more decomposable than words for which the direct route dominates – that is, they are stored together with stronger associations for their component parts. Words for which the direct route tends to dominate have weaker associations with their component parts. Such words are less 'tied' to the semantics of the constituents, more prone to semantic drift, and are rated in experiments as feeling less morphologically complex (Hay 2003). Thus, we can conceive of a word as highly affixed if the component parts play a strong role in lexical access, in which case the word is stored with strong associations with its parts. A word is less affixed if the component parts play a relatively minor role in lexical access, and the word is stored with weak associations with its parts. Less decomposable words fall toward the less creative end of this continuum and are dominated by whole word storage, while more decomposable words fall toward the more creative end of the continuum, and are characterised by a high degree of compositionality.

Having established the notion of degrees of decomposability, it is worth considering some factors which contribute to the degree of decomposability of affixed words. In Hay (2001) I argue that one important factor determining the degree

of decomposability of a morphological form is the relative frequency of a derived word and its base. Words which are more frequent than their bases (e.g. *illegible* is more frequent than *legible*) tend to be less decomposable than words which are less frequent than their bases (*illiberal*, less frequent than *liberal*). *Unfortunate*, for example, is more frequent than *fortunate*, which would tend to facilitate the role of whole word access rather than access via its component parts.

The majority of affixed words in English, unlike *unfortunate,* are less frequent than their bases. The general intuition behind this effect is that the relative salience of the base word within the derived word facilitates the perception of the component constituents. This, in turn, leads to a stronger association between the affixed word and its component parts. Consequently, words which are more frequent than their bases tend to be rated as less 'complex' than words which are less frequent than their bases (Hay 2001), and affixes which are dominated by the latter type of word tend to be more productive overall (Hay and Baayen 2002, 2004).

There is some phonetic evidence for the role of relative frequency in complex words. Hay (2003) demonstrates that words such as *swiftly* (which is more frequent than *swift*) tend to be pronounced with less of a /t/ than words such as *softly* (which is less frequent than *soft*). Note that *swiftly* and *softly* are approximately equal in frequency, it is the relative frequencies of the derived forms and the bases that differ. Words which are less decomposable tend to be associated with a greater degree of reduction at the morphological boundary. In addition, Hay (2003) found that words which are less frequent than their base are more likely to attract a contrastive pitch accent on the prefix – suggesting that the prefix is a transparent and meaning-bearing constituent.

Another factor that seems to facilitate decomposability in English is the phonotactics across the boundary between an affix and a base. It is well established that listeners (even infants) are sensitive to the co-occurrence probabilities of phonemes, and use low probability phoneme transitions to spot the boundaries between words (Saffran et al. 1996a, 1996b). This sensitivity carries over to some degree to affixed words, such that low probability phoneme transitions across morpheme boundaries facilitate parsing and reinforce strong associations between derived forms and their component parts. For example *bowlful* tends to be rated as less decomposed than *pipeful*, because the l][f transition is attested inside monomorphemic words in English (e.g. *dolphin*), whereas the p][f transition is not. Affixes which tend to create many low-probability phoneme transitions tend to be more productive (Hay and Baayen 2004). Prefixed words containing illegal phoneme transitions tend to be more semantically transparent, less polysemous, and are more likely to be less frequent than their base (Hay 2003). Thus, we might expect a word like *uncork*, all other things being equal, to be less decomposable, and stored with weaker associations with its parts than a word like *unhinge* be-

cause the n][k transition is fairly well attested inside English monomorphemic words, but the n][h transition is not.

The general hypothesis that *un* may be longer when it is more meaning-bearing is consistent with work on the length of words, which shows that the informativeness of a word in its local context can predict its duration. More informative words are longer than words which are highly predictable in the context in which they occur (Gregory et al. 1999; Jurafsky et al. 2000).

If we entertain the hypothesis that *un* may be longer in words which have strong associations with their component parts, then two more specific hypotheses will follow. One is that *un* may be shorter in words which are more frequent than their bases, and the other is that *un* may be shorter in words which contain legal phonotactics across the morpheme boundary. If it is the case that the length of *un* correlates with the decomposability of the derived word, then this would provide evidence that gradient morphological structure can be reflected in the fine phonetic detail of spontaneous speech. This would open up a variety of possibilities for future research, including the study of the speech of individuals where a particular word may be formed 'creatively' by one speaker (or on one occasion), while being retrieved whole by another speaker (or by the same speaker on a different occasion). Phonetic evidence on this point would enable us to conduct considerably more sophisticated and detailed work on what exactly constitutes 'lexical creativity' both within individuals and across different groups of speakers.

4.   Methodology

I searched for the orthographic string *un* in the recordings of a subset of speakers from the Origins of New Zealand English (ONZE) Corpora of New Zealand English at the University of Canterbury. The earliest born ONZE speakers were interviewed for a radio broadcast in the 1940s, and the most recent speakers were interviewed by undergraduate students as part of their course requirements. The corpus also includes some interviews which were originally conducted for Oral History projects. A full description of the recordings is available in Gordon et al. (in press), and extensive phonetic analysis of the early recordings has been reported in Gordon et al. (2004). The full set of recordings comprises over 1000 hours of audio. We are in the process of migrating these recordings to a digitally interactive system (Fromont and Hay 2004; Gordon et al. 2005), and the recordings we analysed were those available in our interactive database as of October 2005. From this data, we extracted tokens of *un* in the speech of 244 speakers born in New Zealand between 1857 and 1982. This data-set includes 109 females and 135 males.

We analysed all tokens in which *un* was an affix, and also (for comparison) tokens of *un* in the words *unless* and *until*. These items were chosen for comparison as, unlike other potential candidates such as *uncle* and *under*, they more directly parallel the most common stress pattern for the affixed words (i.e. stress on the base rather than on the prefix). A total of 359 affixed forms and 310 non-affixed forms were analysed.

The length of *un* was measured in milliseconds. Whether the vowel in *un* was full, absent, or reduced to schwa was also recorded. We first compare the phonetics of the affixes and the non-affixes to investigate whether an affixal *un* tends to be longer than a non-affixal *un*, and whether it is less prone to reduction.

5. Results

5.1 Affixes versus non-affixes

In considering the difference between affixed and non-affixed forms, we first examine the likelihood of the vowel being reduced or completely absent. Table 1 shows the distribution, for both affixed and non-affixed forms. Not surprisingly, we find that the *un* in the non-affixed forms is considerably more likely to be reduced. Very few prefixes actually contain a reduced or absent vowel. One of the phonetic qualities that is associated with the prefix *un*, then, is a relatively robustly present full vowel. If the *un-* carries a meaning component, then there is a disincentive to reduce the vowel.

We then considered the relative length of *un*.

Because there are considerable differences in individual speech rate, the comparison of raw length measurements may be misleading. Instead, we attempted to normalise for speech rate to some degree by calculating the length of *un* relative to the length of the following syllable. In defining the following syllable, we assume the maximal onset principle (such that, e.g. the /l/ in *unbalanced* belongs to the following syllable, but the /l/ in *unhealthy* does not). We then fit a linear model designed to predict the relative length of *un*.

Table 1. Comparison of degree of reduction of affixed and non-affixed forms (Comparing full vs not full: X-squared = 328.78, df = 1, p-value < .00001)

	Not-affixed	Affixed
Vowel Absent	95	4
Vowel Reduced	122	7
Full Vowel	93	348

Using the relative length of *un* normalises for speech rate to some degree. However, this also introduces other complicating factors related to inherent differences in the lengths of following syllables. For example, the following syllable may vary in length depending on whether it is stressed, and whether the word occurs utterance finally. In order to inspect reasonably comparable items, we restricted this analysis to cases where *un* did not carry a pitch accent, and where it was produced with a full vowel. Affixed words which did not have lexical stress on the first syllable of the base were also excluded from the analysis. Whether or not the word was utterance final was included as a factor in the statistical model.

Probably the most important factor affecting the length of the following syllable is the weight of that syllable. We therefore coded the following syllable as "light", "heavy" or "superheavy". Light syllables contained a short vowel (i.e. one mora). Heavy syllables contained a long vowel, or a short vowel and a coda (two moras). Superheavy syllables contained at least a long vowel and a coda, or a short vowel and two coda consonants (i.e. three moras). We collapsed *unless* and *until* together into a separate "unaffixed" category in the same factor group.

The details of the linear regression are shown in Tables 2 and 3. Whether or not the word was utterance final had a significant effect. This is because words which are utterance final are prone to phrase-final lengthening (Wightman et al. 1992). This would lengthen the base, and make the relative length of *un* somewhat shorter.

The nature of the following syllable also had a significant effect. The model predictions (i.e. holding utterance-finality constant) are shown in Figure 1. As

Table 2. Wald statistics for model predicting the relative length of un, over both affixed and unaffixed forms

Factor	d.f.	Partial SS	MS	F	p
Syllable Type	3	2.999	0.999	22.2	< .0001
Utterance Final	1	0.745	0.746	16.55	< .0001
REGRESSION	4	3.828	0.957	21.25	< .0001
ERROR	332	14.953	0.045		

Table 3. Coefficients for model predicting the relative length of un over both affixed and unaffixed forms

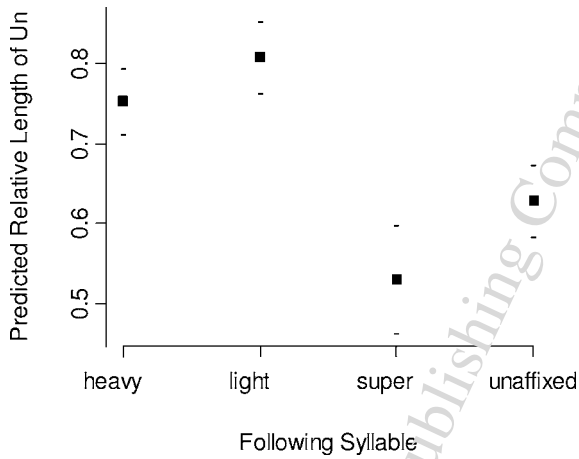|	
Intercept	0.754
Syllable type=light	0.056
Syllable type=super	−0.223
Syllable type=unaffixed	−0.124
Utterance Final=yes	−0.115

Figure 1. Model predictions showing the effect of the following syllable

expected, the heavier the following syllable, the shorter "un" is relative to that syllable. *Unless* and *until* both contain following syllables coded as heavy. However, as can be seen from Figure 1, the *un* in these words is actually considerably (and significantly) shorter than the *un* in the comparable (heavy) affixed forms.

This provides some baseline evidence that when *un* is an affix it tends to be relatively longer than when it is simply a non-meaning-bearing syllable. It will be noted that the words we have used as controls are both function words – an inevitable choice in order to avoid tokens in which the *un* was stressed. Some readers may consider this an unfair comparison, as function words tend to be more reduced than content words (see e.g. Jurafsky et al. 2001; Lavoie 2002; Shi et al. 2005). This argument could, indeed, be rallied against the results regarding the phonetic reduction of the vowel, and so these results should be regarded with some caution. However, in terms of the relative length of *un,* we believe that this should not affect the results. This is because in the function word, we expect the whole word to be prone to shortening so, while the absolute length of *un* may be shorter, the following syllable should also be shorter.

This comparison of affixed forms with *until* and *unless* was intended to establish a simple baseline comparison for our more detailed investigation of affixed forms. When *un* is an affix (as opposed to part of a simple word) there is some evidence that it is less prone to phonetic reduction (either in vowel quality, or through vowel deletion), and it also tends to be longer.

In terms of our wider hypothesis about the phonetics of affixed forms, then, we expect highly creative, highly decompositional words with *un* to have relatively

longer *un*s than words which are highly lexicalised. Following Shi's (2005: 359) argument that the reason function words tend be reduced "can be attributed to their low semantic load and high predictability", we can assume that affixes which carry a smaller semantic load (by virtue of being in highly lexicalised words) should also show more phonetic reduction than affixes which appear in productively coined words and which carry a greater semantic load.

This possibility is explored in the following section.

5.2   The phonetics of affixal un

Only 11 affixes were, in fact, reduced. Obviously this is too small a number to assess what factors may facilitate reduction. It is, however, worth pointing out that the tendency regarding relative frequency goes in the expected direction. As shown in Table 4, while 2.5% of forms which are less frequent than their bases contain reduction of *un*, this is true for 4.3% of the forms which are more frequent than their bases. While by no means statistically conclusive, this is at least consistent with the notion that words which are more frequent than their bases are less likely to be highly decompositional. In such words, the *un* may not function as a distinct meaning-bearing unit, and is thus available for reduction.

As outlined above, in order to assess the effect of length, we measured the length of the affix in milliseconds, relative to the length of the following syllable (assuming maximal onsets). This calculation of the relative length of *un* goes some way toward eliminating speech rate effects. It does, however introduce other potential effects – namely independent variables influencing the length of the following syllable, including its weight, whether it carries lexical stress, whether it is word final, and whether the entire word is phrase final. In an attempt to remove any significant effects of this kind, these were considered in the modelling process. We also considered the part of speech (i.e. distinguishing between verbal and adjectival *un*), whether the base had been recently mentioned in the discourse, and the age and gender of the speaker. The factors of particular interest were whether the base or the full word was more frequent, and the phonotactics of the transition.

Table 4. Comparison of degree of reduction of vowel in affixed forms, by relative frequency of the derived form and the base

	Reduced or Absent	Full Vowel
Base more frequent	6 (2.5%)	234 (97.5%)
Word more frequent	5 (4.3%)	111 (95.7%)

An ordinary least squares model of the relative length of *un* was fit by hand, starting with a relatively saturated model, and removing factors which did not reach significance. The model is shown in Tables 5 and 6. Syllable weight (light, heavy, super-heavy), whether the syllable carried lexical stress, and utterance finality all played a significant role. These effects (from the model – i.e. holding all other significant effects constant) are shown in Figures 2 and 3. As would be expected, factors which would lead to the increased length of the following syllable (a heavier syllable, a lexically stressed syllable, and an utterance final word) all

Table 5. Wald Statistics for Model predicting the relative length of Un in affixed words

Factor	d.f.	Partial SS	MS	F	P
Most frequent (Factor+Higher Order Factors)	2	0.648	0.324	5	0.007
All Interactions	1	0.647	0.647	9.98	0.002
Period (Factor+Higher Order Factors)	2	0.681	0.341	5.26	0.006
All Interactions	1	0.647	0.647	9.98	0.002
Syllable type	2	4.504	2.252	34.77	< .0001
Strong-following	1	1.412	1.412	21.8	< .0001
Phonotactics	3	1.588	0.529	8.16	< .0001
Utterance Final	1	0.324	0.324	5	0.026
Most frequent * Period (Factor+Higher Order Factors)	1	0.647	0.647	9.98	0.002
REGRESSION	10	15.073	1.507	23.27	<.0001
ERROR	334	21.637	0.064		

Table 6. Coefficients for Model predicting the relative length of Un in affixed words

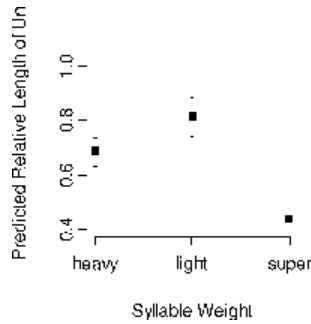|	
Intercept	1.216
Most frequent =word	−0.128
Period=late	−0.082
Syllable type=light	0.129
Syllable type=super	−0.249
Strong-following=yes	−0.249
Phonotactics=good	−0.199
Phonotactics=same	−0.099
Phonotactics=vowel	−0.0001
Utterance Final=yes	−0.071
Most frequent=word * Period=late	0.194

Figure 2. The effect of the following syllable weight on the relative length of 'un'
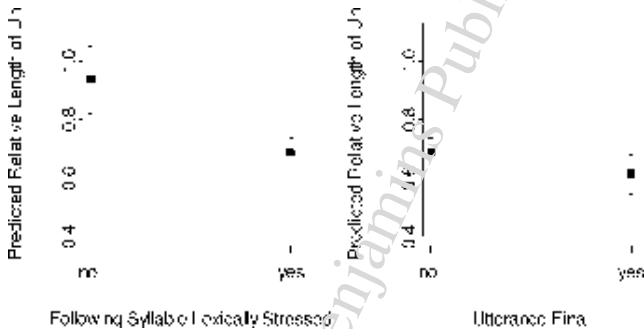

Figure 3. The effect of whether the following syllable carries lexical stress (left panel) and whether the word is utterance final (right panel), upon the relative length of un

decrease the length of *un* relative to the length of the following syllable. It is hoped that the inclusion of such factors in the model effectively subtracts any differences in relative length which are attributable solely due to variation in the length of the following syllable. We hypothesised that any further variation in relative length may be attributable to effects carried by variation in the length of the affix itself. A result which is interesting in this regard is the effect of the phonotactics across the word boundary. This factor contains four categories – the base begins with a vowel (e.g. *uneasy*), with an alveolar nasal (e.g. *unnatural*), with an onset leading to a phonotactically legal transition (e.g. *untidy, unsafe*), and with an onset leading to a phonotactically illegal transition (e.g. *unwell, unpleasant*). Legality was established by checking whether the transition occurs in any monomorphemic words in English by using a corpus of monomorphemes from the CELEX lexical database (see Baayen et al. 1995 for a description of CELEX and Hay et al. 1999 for a description of the subset of monomorphemes). For example the transition n][t occurs inside monomorphemic words (e.g. *winter*), whereas the transition n][w does not.
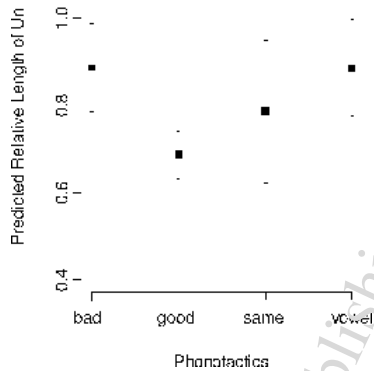
bad=base begins with an onset creating an illegal transition (e.g. unpleasant);
good=base begins with an onset creating a legal transition (e.g. untidy);
same= base begins with alveolar nasal;
vowel= base begins with vowel.

Figure 4. The effect of the phonotactics across the morpheme boundary on the predicted relative length of un

This factor proved to be significant, and the effects are shown in Figure 4. Interestingly the three categories 'bad' (i.e. containing illegal consonantal transition), 'vowel' and 'same' (containing an alveolar nasal) are not significantly different from one another. This is perhaps not surprising as there is a sense in which all of these are illegal. We kept the sequence of nasal-vowel and nasal-n separate as, while these are technically illegal, they do allow for the possibility of resyllabification over the morpheme boundary. That these patterned with the 'bad' consonant-initial bases suggests that this didn't happen. Intriguingly, in words in which the base begins with a consonant that forms a legal transition, the length of the *un* is considerably reduced. This reinforces Hay's (2003) claim that such words are more likely to be accessed whole, and so are more prone to semantic drift. The individual parts are less likely to play a dominant role in such words, relative to the derived form, and so the length of the *un* can be reduced. Note that it is particularly revealing that this result is significant, given that the 'bad' cases are more likely to begin with consonant clusters (e.g. *unpleasant*, *unscramble*). If anything, this should lengthen the following syllable, thereby reducing the relative length of *un*. However, *un* in such cases is still significantly longer, despite this possible bias in the opposite direction.
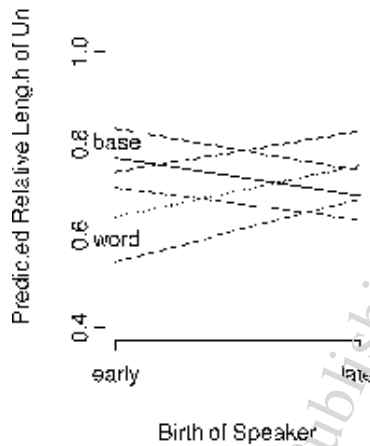
Figure 5. Interaction between the period of birth of a speaker (early = before 1920, late=1920+) and whether the word or the base is more frequent

This result, then, provides evidence in favour of the hypothesis that the relative decomposability of words prefixed in *un-* may be reflected in the length of the prefix.

In order to test for any potential changes over time, the speakers were divided into two groups: 'early' speakers, born before 1920, and 'late' speakers, born in 1920 or later. Recall that I had hypothesised that words where the base is more frequent than the full form are likely to contain a longer *un* than words where the full form is more frequent. This hypothesis was supported, but only for the early speakers. This interaction is shown in Figure 5. For early speakers, as predicted, words which are more frequent than their bases tend to have a relatively shorter (i.e. less affix-like) *un*. For late speakers, this distinction does not hold up (and, if anything, goes slightly in the opposite direction).

This change over time was certainly unexpected – why should only one group show a difference between greater and lesser degrees of lexicalisation?

One possible explanation is that there has been a change in the way that the *un-* prefix is being used in New Zealand English. Affixes certainly come and go in terms of fashion at different times, and can be influenced by a variety of social and contextual factors (see, e.g. Keune et al. 2005; Plag et al. 1999). Perhaps the contemporary New Zealanders are using the prefix less productively. This may explain why there is no distinction between words which are more frequent than their bases and those which are less frequent – perhaps the young New Zealanders are retrieving a greater proportion of their *un-* words whole.

Table 7. Comparison of words containing 'un' and the lexical item 'not', for early and late speakers.(X-squared = 5.9689, df = 1, p-value < 0.02)

	Un	Not
Early	138 (16%)	713 (84%)
Late	218 (13%)	1512 (87%)

The contemporary New Zealanders do seem to be using a slightly smaller range of *un* words. The type/token ratio for the early speakers is .49 (68/138), whereas for the later speakers it is .41 (89/218). This is a small difference, but it may suggest that the proportion of creatively-coined *un-* words being used by the later speakers is somewhat smaller.

Given that the overall amount of speech analysed differed for the early and late speakers, comparing the raw number of *un-* tokens does not help establish whether *un-* words are actually being used more frequently by the earlier speakers. In an attempt to crudely isolate environments where individual speakers could have used an *un-*word but did not, I counted all instances of the word '*not*' occurring in the same data-set. The results of this comparison are shown in Table 7, from which we discover that the earlier speakers are using significantly more *un-* forms (relative to the word '*not*') than the later speakers.

While part of speech was not a significant predictor of prefix length, it is worth separating the two types of *un* here. The data-set collapses two different affixes (verbal, as in *unwind*, and adjectival as in *unwell*). If there has been a change over time, it is worth considering whether this change is carried by one or both of these affixes. The comparison between early and late speakers in Table 7 reaches significance if considered over the verbs alone (3% vs. 1% '*un*', p < .003), but not for adjectives alone, (14% vs. 12% '*un*', p < .13). However, for both groups of words, the later speakers do seem to be producing a lower *un/not* ratio than the early speakers are.

A more direct investigation, in which we analyse the instances in which speakers could have produced an *un-* form but didn't, is difficult to conduct for the verbal items. However, it is possible in the case of adjectives. In order to more directly compare the relevant frequencies for adjectives, I took all occurrences of *not* followed by an item that could be an adjective – that is, 'adjective' was listed as a potential part of speech for that lemma in the CELEX lexical database. I then went through each of these by hand, identifying cases where the meaning could potentially have been conveyed by a prefixed form (e.g. *not pleasant, not healthy, not sure…*). There were 50 such examples amongst the early speakers, and 135 examples produced by the late speakers. The late speakers, therefore, have a sig-

Table 8. Comparison of 'un'-prefixed adjectival forms, and sequences of not+adj which could be expressed with a prefixed form. (X-squared= 4.23 ,df=1, p < .04)

	Un+adj	Not+Adj
Early	114 (70%)	50 (30%)
Late	199 (60%)	135 (40%)

nificantly lower ratio of *Un+adj* versus *Not+Adj* than the early speakers do. This distribution is shown in Table 8.

This analysis presents further (and more direct) evidence that the later speakers may be using adjectival *un* less productively than earlier speakers. They have a lower type/token ratio, at least for adjectives, and adjectives also have a lower "*un/not*" ratio.

While it is harder to find direct evidence with the verbs, it does seem that the use of *un* with verbs may also be decreasing. The earlier speakers actually use more verbs relative to adjectives than the later speakers (17% of the early speakers' tokens are verbs, and only 9% of the late speakers', X-squared = 5.2003, df = 1, p < .03). However, the overall number of verbal tokens is small and, unlike the adjectival tokens, there is no straightforward alternate phrasing that one could easily search for.

6.   Discussion and conclusion

This study has provided evidence that the gradient decompositionality of affixed forms is reflected in phonetic detail of spontaneous speech. The particular focus of this paper was the length of *un*, and we found that factors which tend to facilitate decomposability are also associated with longer prefixes. Highly creative, productively coined *un-* words, then, are likely to reveal this productivity in the relative length of the prefix.

In addition, the study also reveals an intriguing change over time, in which contemporary New Zealanders seem to be using *un-* less productively than their counterparts born in the 19th and early 20th Century. This is interesting, particularly as we first discovered this shift by observing a subtle change over time with regard to the importance of relative frequency in predicting the length of *un*. More detailed probing revealed that contemporary New Zealanders tend to use *un-* slightly less frequently than older New Zealanders, and are more likely to use sequences of *not*+adjective.

This suggests that attention to the phonetic detail of affixed words may be a valuable source of information in the study of morphological creativity, and may

potentially enable us to assess varying degrees of compositionality in productive word-formation in the speech of specific individuals on specific occasions.

Hohenhaus (this volume) argues that there is a cline between lexical creativity, on the one hand, and productivity on the other hand (where productivity is highly rule-governed, and lexical creativity is not). This certainly seems to be a continuum, and one could also argue that the cline extends further to include monomorphemic words. There is a continuum between words which contain no identifiable subparts, through phonaesthemes (such as the *fl* in words relating to liquid such as *flow* and *float* – see Bergen 2004), through affixed words (showing different degrees of productivity), right through to highly creative nonce words which have been consciously coined.

If we conceptualise this continuum from a cognitive perspective – reflecting the varying degrees of compositionality that a speaker engages in to produce a word – then merely observing the existence or use of a word can not help us position that word with any accuracy on this continuum. This is because the degree of composition involved for a given speaker may reflect a number of factors that vary across speakers, across social contexts, and over time. The degree of composition involved may reflect whether (or how often) the relevant word or its component parts have previously been encountered, if at all. It may also reflect whether the component parts were uttered in the preceding discourse. Finally, it may depend on whether the particular word formation strategy being used is a highly productive one for that era, and/or that speaker. For example, blends such as those discussed by Lehrer (this volume), may reflect high levels of creativity by the speaker who first coins them. But for a listener who hears the blend and repeats it on a later occasion, this second usage is already somewhat less creative and less compositional than the first occurrence. In time, blends such as *motel* or email may function as effectively monomorphemic for many speakers. For speakers of early New Zealand English, a word like *unbiased* may once have been highly compositional whereas for a speaker of contemporary New Zealand English, the word would be positioned closer to the monomorphemic end of the continuum.

The case study presented in this paper has revealed that degrees of decompositionality may be reflected in phonetic detail. Thus, phonetics may prove to provide an insight into the nature of the continuum between stored words and highly creative or nonce words. Specific words may be differently positioned on this continuum for different speakers, or on different occasions. By examining the phonetic detail of morphologically complex words, we may be able to make some progress in understanding the highly dynamic and changing nature of lexical creativity, and the ways in which it varies across speakers, contexts and time.

References

Baayen, R. Harald, Richard Piepenbrock and L. Gulikers 1995. The CELEX Lexical Database (release 2) cd-rom. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania (Distributor).

Baayen, R. Harald 1992. Quantitative aspects of Morphological Productivity, in G. Booij and J. van Marle (eds), *Yearbook of Morphology* 1991, 109–150 Kluwer Academic Publishers, Dordrecht.

Baayen, R. Harald and R. Schreuder 1999. War and Peace: Morphemes and Full Forms in a Non-interactive activation Parallel Dual-Route Model. *Brain and Language* 68, 27–32.

Bergen, B. J. 2004 "The psychological reality of phonaesthemes." *Language* 80: 290–311.

Fromont, Robert and Jennifer Hay 2004. "Development of an Interactive Digital Corpus of New Zealand English." Paper presented at New Zealand Language and Society Conference, Palmerston North, September 2004.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury and Peter Trudgill 2004. *New Zealand English*: *Its Origins and Evolution*. Cambridge University Press, Cambridge.

Gordon, Elizabeth, Margaret Maclagan and Jennifer Hay (in press). "The ONZE Corpus." In Beal, J. C, Corrigan, K. P, Moisl, H. (eds) Using Unconventional Digital Language Corpora *Volume 2: Diachronic Corpora*. Palgrave.

Gregory, Michelle, William Raymond, Alan Bell, Eric Fosler-Lussier and Daniel Jurafsky (eds.)1999. "The effects of collocational strength and contextual predictability in lexical production" *Proceedings of the Chicago Linguistics Society* 35. 151–166.

Guy, Gregory R. 1980. "Variation in the Group and the Individual: The Case of Final Stop Deletion." In *Locating Language in Time and Space,* ed. W. Labov, 1–36. New York: Academic Press.

Guy, Gregory 1991. "Explanation in Variable Phonology: An Exponential Model of Morphological Constraints." *Language Variation and Change* 3: 1–22.

Hay, Jennifer 2003. *Causes and Consequences of Word Structure.* New York and London: Routledge.

Hay, Jennifer 2001. "Lexical Frequency in Morphology: Is Everything Relative?" *Linguistics*, 39 (6), 2001, pg 1041–1070.

Hay, Jennifer and Baayen, Harald 2005. "Shifting paradigms: gradient structure in morphology." In *Trends in Cognitive Sciences,* 9(7), 342–348.

Hay , Jennifer and Harald Baayen 2004. "Phonotactics, Parsing and Productivity." *Italian Journal of Linguistics*, 15(1), 99–130.

Hay, Jennifer and Harald Baayen 2002. "Parsing and Productivity." In Booij, G. and van Marle, J. (eds.) *Yearbook of Morphology 2001*. 203–235 Kluwer Academic Publishers.

Hay, Jennifer, Janet Pierrehumbert and Mary Beckman 2004. "Speech Perception, Well-formedness and the Statistics of the Lexicon." In Local, J., Ogden, R. and Temple, R (eds) *Phonetic Interpretation: Papers in Laboratory Phonology VI*. 58–74 Cambridge University Press: Cambridge.

Jurafsky, Daniel, Alan Bell, Michelle Gregory and William Raymond 2001. "Probabilistic relations between words: Evidence from reduction in lexical production." In J. Bybee and P. Hopper (eds) *Frequency and the Emergence of Linguistic Structure*. 229–254 Amsterdam: John Benjamins.

Keune, K., M. Ernestus, R. Van Hout and R. Baayen 2005. "Social, Geographical, and Register Variation in Dutch: From Written MOGELIJK to Spoken MOK." *Corpus Linguistics and Linguistic Theory. In press.*

Labov, Willam 1989. "The Child as Linguistic Historian." *Language Variation and Change* 1: 85–97.

Lavoie, Lisa 2002. "Some influences on the realization of for and four in American English." *JIPA* 32: 175–202.

Plag, Ingo, Christine Dalton-Puffer and Harald Baayen 1999. "Productivity and register." *Journal of English Language and Linguistics* 3, 209–228.

Saffran, J. R., R. N Aslin and E. L. Newport 1996a. "Statistical learning by 8-month old infants." *Science* 274, 1926–1928.

Saffran, J. R., Newport, E. L. and Aslin, R. N. 1996b. "Word Segmentation: The Role of Distributional Cues." *Journal of Memory and Language* 35, 606–621.

Shi, Rushen, Bryan Gick, Dara Kanwischer and Ian Wilson 2005. Frequency and Category Factors in the Reduction and Assimilation of Function Words: EPG and Acoustic Measures." *Journal of Psycholinguistics Research* 34, 4, 341–364.

Wightman, Colin W., Shattuck-Hufnagel, Stefanie and Ostendorf, Mari 1992. "Segment durations in the vicinity of prosodic phrase boundaries. JASA 91(3), 1707–1717.