Morphological Approaches for an English Pronunciation Lexicon

Susan Fitt

Centre for Speech Technology Research
University of Edinburgh
s.fitt@ed.ac.uk

Abstract

Most pronunciation lexica for speech synthesis in English take no account of morphology. Here we demonstrate the benefits of including a morphological breakdown in the transcription. These include maintaining consistency, developing the symbol set and providing the environmental description for allophones and phonetic variables. Our approach does not use a full morphological generator, but includes morphological boundaries in the lexicon.

1. Introduction

Morphology is rarely addressed in speech technology. It is, however, of obvious benefit for some languages, such as German [1]. German has a high number of words formed by agglutination; speakers easily produce new formations, and it is very difficult for a lexicon to cover all the possibilities. A morphological component which extends a German lexicon is of great benefit.

In English we can also make new and understandable creations, ranging from the useful, such as 'formability', to the preposterous or humorous, e.g. 'understandification'. However, these form only a small part of a speaker's output. Furthermore, the accurate derivation of existing words from roots is complex. This is presumably why development of a morphological generator is generally regarded as low priority.

Here we will present the benefits of using morphology in a speech synthesis lexicon, and show how most of these can be gained from a compromise solution – including a morphological breakdown in the lexicon.

2. Background

The lexicon described here is an accent-independent pronunciation lexicon of English, designed to facilitate the synthesis of regional accents. The basics of the lexicon are described in other papers (see website for up-to-date papers, or [2] for a description of an earlier version of the lexicon). The lexicon uses keysymbols, a kind of meta-phoneme, to encode pronunciation differences across English accents, so numerous accents can be synthesised using a single lexicon.

There are a number of features of this lexicon which made the inclusion of morphology in some form especially useful. We discuss the benefits that would arise from both a full morphological generator, and morphological annotation in the lexicon, and explain how they are particularly advantageous for our accent-independent dictionary.

2.1. Generation of new words

A full morphological generator has the obvious benefit of simplifying the addition of new words. We would be able to give transcriptions for nonce-words such as 'formability'. We could even choose to use the adjectival creation

'apply' { * a . p l }.> iy >

, generated from 'apple' + 'y', rather than the usual verb entry

'apply' {@ . p l * ae }

(See Table 1 for descriptions of morpheme markers in these examples.)

We would also gain in transcription accuracy for new words. The accent-independent lexicon is more complex than most as it needs to contain more information in order to cover numerous accents. We have a larger symbol set than usual, consisting of basic symbols and a set of typographical conventions which extend the basic symbol set, for example square brackets represent a deletable segment. The transcription for 'herb', { [h] * er r b }, thus contains an |h| which is present in UK accents (or at least those which do not use h-dropping) but not present in US accents.

This complexity makes the automatic generation of new derivations or compounds especially appealing, since the more complex the transcription, the more likely the errors when adding new words by hand.

2.2. Consistency of pronunciation

A related topic is the consistency of transcriptions for common words which are usually contained in the lexicon. For example, the lexicon contains numerous 'o'-type vowels, representing the vowels in NORTH, FORCE, THOUGHT, and so on (c.f. [3]). The vowel in 'horse' belongs to the NORTH set; the lexeme 'horse' occurs in no less than 105 words in our current lexicon of 118,000 entries, ranging from the obvious 'horses' and 'horsey' to 'horselaughs' and 'stockhorse'. Being able to identify one root entry for 'horse' and relating the others to these makes it easier to maintain consistency. A morphological generator would be the most accurate way of doing this; morphological annotation, however, does aid the process significantly, as we will show below.

2.3. Development of keysymbols

As the lexicon covers multiple accents, it is open to revision when new accents are added. For instance, the long and short |a| described by Fudge [4], which differentiate 'jam' and 'sham', is not included as we consider it to be of minor importance. However, if we were to synthesise Fudge's accent, we would need to transcribe this split. It is much easier to add a new symbol if the lexicon is small and if we only need to change each lexeme once. Morphological information gives both of these benefits.

2.4. Description of exceptions

Exceptions are also easier to state if we only need to list them once. This simplifies the listing, and makes the lexicon system easier to maintain. For example, we transcribe 'iron' as { * ae @r r n }, but for Scotland we need to make this an exception, { * ae . r @ n }. If we have a way of generating derivations, we do not need to list 'irons', 'ironing' and so on as exceptions, but can generate them as needed; we can also, of course, generate new words such as 'ironability', all based on the Scottish root { * ae . r @ n }.

2.5. Allophones and other pronunciation rules

Despite all these benefits, the crucial factor in deciding to include morphology was allophones. For example, in Belfast there is a contrast between dental |d|, |t| in the monomorphemic 'spider', 'matter', and nondental |d|, |t|, in 'wider', 'fatter', where they precede a free morpheme boundary [5]. We cannot transcribe allophones in the lexicon as they vary too much across accents, but these examples cannot be derived unless our allophone rules have access to morphological boundaries.

Another pronunciation rule which is easier to state given morphological information is "-ing" reduction. "-ing", usually pronounced as |i ng|, can be pronounced as |i n| or syllabic |n|, but only under certain conditions: it must be unstressed and must be the final string in a multisyllabic free root or suffix. So, for example, we can reduce the |i ng| in 'pudding', and also in 'puddings', but not in 'sing' or 'singer', where it is stressed and is a monosyllabic root. Importantly, we cannot reduce the |i ng| of 'batwing' although it is usually considered to be unstressed, since the 'wing' root is a monosyllable; so, morphology helps to block reduction in this case. (This analysis is slightly simplified due to space; 'something' and 'anything' potentially contain a monosyllabic root 'thing', which is reducible and complicates the rule.)

3. Automatic generation and decomposition

We began optimistically, with the aim of creating a full-blown morphological component which would enable us to store the pronunciations of roots and affixes.

As with any grand idea, there are a number of difficulties. Firstly, storing roots and affixes alone does not leave us any way of storing related information such as word frequency. Secondly, if we generate every lexical entry at run time the processing is slow, acceptable for nonce-words but not for common words. If, on the other hand, we generate all possible words to create a lexicon, the lexicon will be huge and will contain a large number of very unlikely words, which is also inefficient, and for speech recognition is likely to lead to a high error rate.

The proposed solution (see [2]) was to have two types of lexica: pronunciation and orthographic. Roots and affixes, as well as irregular derivations, would be listed in pronunciation lexica; spellings, frequency, and part of speech of derived headwords would be listed in the orthographic lexicon. Prior to synthesis we would generate a lexicon using the morphological component to combine the roots and affixes; only forms which matched the entries in the orthographic dictionary would be included in our output lexicon. The morphological component would still be available to generate words not found in the lexicon. So, we would list 'apply' (verb) in the orthographic lexicon, and this would be given the pronunciation { @ . p l * ae }. When we came to synthesise 'apply', the output lexicon would give us the default { @ . p l * ae }. If, however, we specified that we required an adjective, the morphological component would be able to produce { * a . p l }.> iy >.

At first all went well, and plurals, gerunds and the like were freely produced from root forms. For many of the basic word categories it is straightforward to produce both orthographic and pronunciation derivations. For example, in the case of plural nouns we need simple orthographic adjustment rules for final "y" ('army'-'armies'), and we need pronunciation rules to specify that the plural suffix |. I7 z| converts to |z| after voiced stops and |s| after voiceless stops, and so on. However, there were further problems.

3.1. Suffix combinations/identification

There are two potential approaches for a morphological component. One is the creation of new forms using the base elements, without reference to a target orthography. The second is morphological decomposition.

The first, free morphological generation, is obviously useful for producing new words to add to a lexicon, and for checking, for example, that all suitable derivations have been included. The difficulty with this approach is that it is not simple to specify which affixes may co-occur, and what order they should appear in. Mohanan suggests that some of these cases can be solved by splitting an affix which behaves in two different manners into two separate affixes. For example '-ment' in 'governmental' precedes what he terms a Class I affix, as it precedes '-al', another Class I affix, while '-ment' in fulfillment' is Class II and occurs later in the affixation process ([6], p. 50).

Roots and affixes cannot always co-occur either. For example, '-ity' and '-ness' often attach to the same roots, ('uniformity', 'uniformness', 'obesity', 'obeseness'), but some combinations are not possible: 'abruptness', but not 'abruptity'; 'mentality' but not 'mentalness'. Some such restrictions are a result of the linguistic origin of the word: for instance, singular nouns of Latin origin ending in '-us' have plural forms with '-i', e.g. 'cactus', 'cacti'; this pattern occurs in a number of English words. However, it does not apply to '-us' words of other origins, such as the Dutch 'walrus', so to produce valid output we need etymological information. To further complicate the matter, there are also a few words of Latin origin which do not follow this rule, e.g. 'bonus', 'omnibus'. Of course, there may be occasions when speakers combine incompatible forms, either for effect or through lack of knowledge, and we would then need to defy our constraints in order to generate a pronunciation.

These difficulties suggested a preference for the second approach, i.e. to decompose existing orthographies. This would be needed in any case for analysing words not listed in the orthographic lexicon. It avoids many of the complications of free generation, as we do not need to define a hierarchy of constraints and preferences for co-occurrence, but instead can simply identify what is presented.

Of course, errors may occur in decomposition, but for the most part these erroneous analyses should be discounted as they will be overruled by complete roots found in the pronunciation lexicon, or by checking the part of speech in the orthographic lexicon. For example, 'apply' would be listed as a verb, and so would not be analysed as the adjectival formation 'apple' + 'y'; the complete root 'relay' would override the breakdown 're' + 'lay'. But, while 'apple' + 'y' is rare, 're' + 'lay' is not so rare, and 'mane' + 's' ("horsehair") is more frequent than 'manes' ("spirits of the dead"), which has the same category of plural noun and would be have to be a root entry.

3.2. Stress

The most successful decompositions were on compound words. This is not surprising, since they are not generally subject to orthographic adjustment and so are easier to break down into morphemes. We developed a program which analysed compounds into roots found in the lexicon, matched the categories of the roots against categories in the lexicon, and compared these to permissible combinations, for instance adjective-noun as in 'hotdog'. Where more than one analysis was possible, the analysis whose roots had the highest combined frequency was ranked highest, except for single letter morphemes which tend to have high frequency but low usage in compounds. So, 'buttonhole' was correctly analysed as 'button-hole', noun-noun, rather than 'but-ton-hole', conjunction-noun-noun, and 'carphone' was analysed as 'car-phone', root frequencies 122606 and 63102, rather than 'carp-hone', root frequencies 617 and 111. 'Email' was analysed as the single letter 'e' + 'mail' since there was no competing analysis.

However, stress proved to be problematic. While stress on compounds is generally predictable according to part-of-speech of the roots and of the whole, there are exceptions. For example, adjective+noun=noun usually results in stress on the first element, as in 'hotdog'. 'Goodwill', on the other hand, has the stress on the second element. Although the decomposition was very accurate, stress errors occurred in a number of the output pronunciations.

3.3. Diminishing returns

As with many areas of both lexicography and speech technology, morphological analysis is subject to diminishing returns. As noted earlier, simple, common categories such as regular plurals are easy to decompose and can be assigned an accurate pronunciation. As we move into more complex categories, we start to write ever more complex rules to account for smaller and smaller groups of data. Exceptions also become an increasing problem. While an automatic morphological decomposition and generation is of obvious benefit in producing new words, the benefit in terms of the existing lexicon was not as great.

4. Annotating morphology in the lexicon

Due to the other requirements of the lexicon, particularly allophones, we still needed a morphological breakdown. At this point we turned to a compromise solution: annotating morphological boundaries in the lexicon. This allows us to use semi-automatic methods rather than the fully automatic methods described above. We can use the automatic methods to produce an analysis of existing words, and hand-edit to correct errors and allow for exceptions.

The other considerations noted above (consistency of pronunciation, development of keysymbols, description of exceptions and production of allophones) helped to establish the priorities in choosing which boundaries to annotate and the symbols to use.

4.1. Boundaries

The allophone and variable rules we have come across so far all depend on free morpheme boundaries or suffix boundaries, which can be considered as free unit boundaries, i.e. they can form the end of a word. For instance, Scottish Vowel Length, which dictates that 'agree' + 'ed' is different from 'greed', is conditioned by the free morpheme boundary of 'agree'. The "-ing" reduction rule applies either at a free morpheme boundary, as in 'pudding', or a suffix boundary, as in 'waiting', but not to a monosyllabic free morpheme. Therefore, these boundaries are of primary importance.

Further development of keysymbols, and maintaining consistency, favour annotating bound morphemes as well as free ones. For example, the verb ending '-ise' can attach to free roots such as 'victim', giving 'victimise', but it also forms part of words such as 'utilise', 'memorise', whose stems also form parts of paradigms ('utility', 'memorial' and so on). Including a marker at the internal boundaries in these words helps us both to identify the component parts, for comparison with the components in other words, and to distinguish the word from other roots; for example if we split 'moderate' into 'moder' + 'ate' we are implicitly linking it with 'moderacy' and '-ate', and ruling out any link with 'mode' or 'rate'.

4.2. Symbols

Having decided which boundaries to annotate, we needed a symbol scheme. This should allow easy identification of the important boundaries, and should be legible and consistent.

The clearest scheme that we tested involved marking the morphemes rather than marking the boundaries. This means that rather than using a single symbol to mark the boundary, for instance 'agreed'

@ . g r * ii + d

we use a marker at each side of each morpheme, for example

{ @ . g r * ii }> d >

Markers	Meaning	Example morpheme
{}	free root	{agree}
<<	prefix	<de<
>>	suffix	>ing>
##	word (for concatenating lexical entries)	#this##is##a##string#
==	internal boundary	{moder==ate}
$	variant pronunciation of free root	acidic { @ . s * i d $}.> i k >

Table 1: Morpheme boundary markers

The first four boundaries in Table 1 all surround the morphemes they annotate. This, combined with the choice of brackets, enables easy identification of the boundaries which are important to us. For example, free units will always be surrounded by outward facing brackets. So, in a compound word we can identify the component words by identifying opposing brackets, for instance in 'sleepyhead' '{sleep}>y>{head}' the main boundary is at the opposing >{, giving us 'sleepy' + 'head', rather than at }>, which face the same way. The boundaries are both visually logical and easy to specify when we come to write rules. Bound morpheme markers do not surround the morpheme; in the schemes we tried, such as '=moder=>ate>', concatenation of multiple morphemes leads to difficulty in identifying the primary components.

4.3. Generating boundaries

4.3.1. Pronunciation field

The boundaries on the pronunciation field were generated semi-automatically, as described earlier. Part-of-speech information, comparison with other morphemes, and adjustment rules were used to produce decompositions, which were hand checked. Uncommon analyses and internal boundaries were mostly produced by hand.

Morphemes were treated as free roots if they were either exactly the same as the free-standing root, or if they differed in certain predictable ways, such as stress shifting; the latter were annotated with a dollar sign. An example pronunciation field is 'oversimplify':

< ~ ou . v @r r <.{ s * i m . p l }> I2 . f ae >

4.3.2. Enriched orthography field

This consists of the orthography annotated with the same morphemes as the pronunciation field, for example

<over<{simple}>ify>

This was generated automatically from the markers on the pronunciation field using a matching algorithm. Firstly a segmental match was used to line up the graphemes and pronunciation symbols. Then the resulting breakdown was compared to existing free roots in the lexicon, both orthographic and pronunciation, and to orthographic adjustment rules. So, in the above example {simpl} was altered to {simple}. Affixes were also adjusted in some cases, for example 'oversimplifies' becomes

<over<{simple}>ify>>s>

This algorithm produced a high degree of accuracy, although it did result in a few errors, for example 'humanity' was analysed as 'humane' + 'ity' rather than 'human' + 'ity'.

4.4. Using the boundaries

The lexicon has been annotated with morpheme boundaries and we are able to use them productively.

4.4.1. Allophones and other pronunciation rules

To take our earlier examples, the morpheme boundary in '{agree}>d>' enables us to trigger the Scottish Vowel length rule; we can also specify the environment for "-ing" reduction more easily and accurately than without the aid of morphemes. Another example is t-glottalisation. In most accents this cannot occur at the start of a free root. The boundaries enable us to transform the second |t| in 'potato':

{ p @ . t * ei . t ou }, { p @ . t * ei . ? ou }

but block the rule for the first |t| in 'atonality':

< ~ ee <.{ t ou n }.> * a l >.> @ . t iy >,
< ~ ee <.{ t ou n }.> * a l >.> @ . ? iy >

4.4.2. Exceptions

We deal with exceptions by listing only the roots in the exceptions list (except for a very few instances where the derivation is an exception but not the root). We then use a program to match the enriched orthography and the original pronunciation. So, the Scottish 'iron' is listed just once, and wherever we find the combination of enriched orthography '{iron}' and pronunciation { * ae @r r n }, or its destressed counterparts, they are replaced with { * ae . r @ n }.

4.4.3. Keysymbol usage

The boundaries are also used in checking consistency and adding new keysymbols. For example, a late addition to our keysymbol set was the distinction between |ei| in WAIST and |ee| in WASTE, a distinction made in, for instance, some Welsh accents [7]. This is closely linked to orthography, with digraphs such as "ai" generally using |ei| and other orthographic forms such as "a" generally pronounced |e|. Use of the morphological breakdown enabled easier identification of segments which fitted the criteria, and also enabled cross-checking of morphemes. This made the symbol split much easier and quicker.

5. Conclusions

A complete morphological component is a nice idea, but the disadvantages of complex rules, inaccurate derivations and inaccurate pronunciations mean that orthographic decomposition and phonetic re-generation is not the best solution for providing core lexical entries.

On the other hand, morphological annotation in the lexicon provides most of the advantages of the decomposition/regeneration approach without the disadvantages. The breakdown transcribed in the lexicon enables us to specify environments for pronunciation rules, simplify exceptions listings, maintain consistency and simplify development of the transcriptions.

6. References

[1] Mengel, Andreas (1999). A phonetic morpheme lexicon for German. *Proceedings: ICPhS 99.*

[2] Fitt, Susan, and Isard, Stephen (1999). Synthesis of regional English using a keyword lexicon. *Proceedings: Eurospeech 99*. Vol. 2, pp. 823-6.

[3] Wells, John C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

[4] Fudge, Erik C. (1977). Long and short [ae] in one Southern British speaker's English. *Journal of the International Phonetic Association*, Vol. 7, pp. 55-65.

[5] Kaisse, Ellen M., and Hargus, Sharon (1994). When do linked structures evade structure preservation? In: Richard Wiese (ed.), *Recent developments in Lexical Phonology*, pp. 185-204.

[6] Mohanan, Karuvannur Puthanveettil (1982). *Lexical Phonology*. Dordrecht: Reidel.

[7] Tench, Paul (1990). The pronunciation of English in Abercrave. In: Nikolas Coupland (ed.), *English in Wales: diversity, conflict and change*, pp. 130-41. Clevedon: Multilingual Matters.