Proceedings of The Institute of Acoustics

Lexical stress assignment in English: a metrical approach

B J Williams

Speech Research. IBM UK Scientific Centre,
Athelstan House, St Clement Street, WINCHESTER, SO23 9DR

Introduction

The placement of stress in English words has sometimes been viewed as a purely lexical phenomenon, stored as part of the lexical entry for each word. If this were true, it would not be possible to synthesise English from completely unrestricted text, as a dictionary-based system would be limited to the words or morphemes already stored. However, English lexical stress can be predicted by rule, given the correct information. The following is an outline of one currently-implemented system for determining the placement of primary lexical stress, and for making stress-determined modifications to the input string. This system is based on hierarchical word-level units framed within the terms of metrical phonology, but taking into account a wider domain than phonology alone.

The system accepts as input a string of phonemes containing syllable and morpheme boundaries, together with the information noun or non-noun. Its output is the phonemic string marked for primary and secondary stress, with any shortened or reduced unstressed vowels also marked. A simple case is the word *Hannibal*, which contains no morpheme boundaries. It would be marked as a noun and given the input phonemic transcription /hæ-nɪ-bæl/, with syllable boundaries added. Note that no vowel reduction has been marked at this stage. The output would be the string /hæ'-n[ɪ]-bl̩/, with primary stress on the first syllable, a syllabic consonant replacing the vowel of the unstressed final syllable, and the unstressed /ɪ/ of the second syllable marked for a centralised vowel quality.

The stress assignment program has been written as part of a larger project involving the synthesis of English speech by rule from unrestricted input text. Thus it assumes input from the letter-to-phoneme component of the system, and its output will form part of the input to the synthesis algorithm.

Characteristics of the input string

First of all, the morphological boundaries are inserted by a preceding morph-stripping algorithm based on a look-up table of affixes. This has not yet been implemented, but it is hoped that storage of affix morphs only will be needed. Affixes are divided into two classes, here described as strong and weak affixes. To the strong class belong those affixes, such as *-ation*, *-ic*, which affect word stress placement, while to the other class belong affixes, such as *-ed*, *be-*, *-al*, which have no such effect (see [1,8] for a fuller discussion). This distinction is essential for the correct parsing of the input string into metrical foot units. For example, the word *consideration* would be divided as follows: con< <sider>ation - where < < marks a *weak prefix boundary* and > marks a *strong suffix boundary*.

The initial phonemic transcription is provided by a set of letter-to-phoneme rules, augmented by an exceptions list [2]. The rules, it is hoped, will not need to specify reduced vowels such as schwa. Since the reduced vowels, being stress-determined, cannot be inferred directly from the orthography, any attempt to do so will involve handling the stress twice in some form or other - with obvious losses in efficiency and generality. Clearly, it is more economical to allow default full vowels in the initial 'phoneme' string, with subsequent stress-conditioned reduction. For example, the initial phonemic form of the word *consideration* would be /kɒn< <sɪdɛr>eɪʃn̩/.

Lexical stress assignment in English: a metrical approach

The syllabification algorithm is based on the method put forward by O'Connor and Trim [4] and referred to by Selkirk [6] as the 'Maximal Syllable Onset Principle'. According to this principle, syllable division is carried out according to the permitted initial and final consonant combinations of English. If a choice should then arise for a particular boundary, the solution is to assign as many consonants as possible to the start of a following syllable. The word *consideration* would thus be syllabified as /kɒn-< <sɪ-dɛ-r>eɪ-ʃn̩/.

Certain part-of-speech information must also be supplied for the correct operation of a rule similar to the Lexical Category Prominence Rule (see below). This information, which probably need be no more than the distinction noun or non-noun, will be supplied by a syntactic parsing program. For example, the word *consideration* would be specified as a noun.

Processing of the input string

The input phonemes are first converted into a representation of their segment-type, using a look-up table. There are four categories of segment-type: consonant (C), long vowel (V), short vowel (v), and reduced vowel or syllabic consonant (E). This process is similar to the vowel-consonant mapping of Segre et al. [5] but includes a finer classification of the vowels. The process is necessary for a correct assignment of 'foot units' at a later stage. The segment-type representation of the word *consideration* would be: *CE-< <Cv-Cv-C>V-CE.*

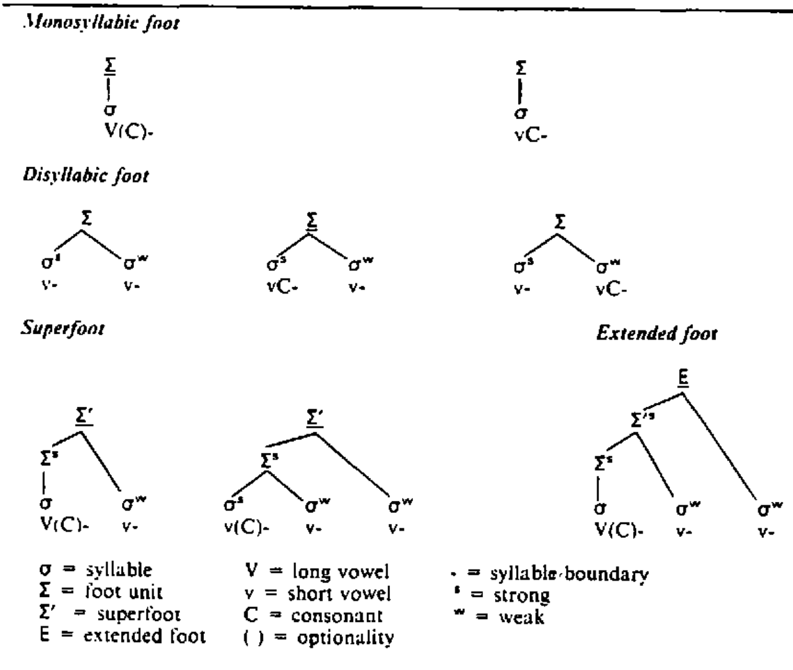

Figure 1. Basic stress foot units

Lexical stress assignment in English: a metrical approach

The morphological boundaries are then made use of to determine the stress domain. Weak affixes, such as *a-*, *con-*, or *-es*, are ignored and form no part of the stress domain, while strong affixes, such as *over-* and *-graphy* are included with the root morph in the stress domain string forming the input to the next stage. The stress domain for *consideration* would be the string *Cv-Cv-C > V-CE* corresponding to *-sideration*.

The next stage parses the stress domain into stress foot units. These units, first described by Selkirk [6,7], are hierarchical phonological units between the syllable and word levels, within which syllables are grouped together in strong or weak relations according to such considerations as vowel length and number of following consonants. A few modifications have been made to Selkirk's original stress foot templates, such that the set of possible stress foot units is made up of the structures shown in figure 1.

Those structures whose topmost nodes are underlined are candidates for anacrusis, ie. a preceding unstressed syllable comprising a short vowel with no syllable-final consonants. It is claimed that all stress domains in English can be parsed in terms of only this set of basic foot units. The representation of the word *consideration* in terms of stress foot units is seen in figure 2.

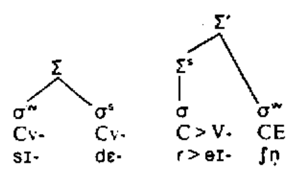

Figure 2. Stress foot units for *consideration*

Once the syllables of the stress domain have been bracketed together into foot units, these units are compared in order to determine strong and weak units. This is done according to a rule based on the Lexical Category Prominence Rule of Liberman and Prince [3]. The rule states that, in a noun, given two nodes, the second is strong if and only if it branches. In a verb, the second node of the two is always strong. Starting from the end of the stress domain, the strong foot unit is determined; if the word is long, the next strongest foot unit may also be considered, for purposes of secondary stress assignment. It seems necessary to make only the simple distinction noun or non-noun for this purpose. The full stress foot bracketing of the word *consideration* is shown in figure 3.

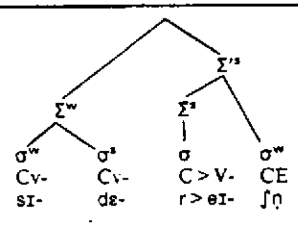

Figure 3. Full stress foot bracketing for *consideration*

The string is then converted back to the original phonemic representation and loses the morpheme boundaries, though retaining the foot unit boundaries. The next stage is that of vowel reduction, which takes different forms according to the type of foot unit and length of vowel.

Lexical stress assignment in English: a metrical approach

In monosyllabic feet with a long vowel. /i, eɪ, aɪ/ become /ɪ/. /u/ becomes /ʊ/, and /ɜ/ becomes /ə/ In monosyllabic feet with a short vowel followed by /l/, /m/ or /n/ within the same syllable, the vowel is deleted and the consonant becomes syllabic. In a weak syllable not constituting a foot in its own right, /æ, ɒ, ʌ, ɜ/ become /ə/, and /ɪ, ʊ/ become marked as more central in quality. If /l/, /m/, or /n/ follow within the same syllable, the vowel is deleted and the consonant becomes syllabic. The final form of the word *consideration* is /kn̩-sɪ,-də-reɪ'-ʃn̩/. The primary stress falls on the fourth syllable, and secondary stress on the second syllable, while the vowel of the third syllable has been reduced to schwa.

The output string thus contains the phonemic transcription with syllable boundaries, marked for primary and secondary stress, with appropriate vowels reduced or converted to sylabic consonants. This output string will form the input to the phoneme-to-allophone rules in the overall synthesis system.

Advantages of the process

If lexical stress can be specified in this way with reasonable accuracy, it would relieve the letter-to-phoneme rules of a task which they can only carry out by specifying the stress an additional time. Many vowel graphemes would then default to the standard - eg. /æ/, /ɛ/, for orthographic *a* and *e* respectively. The question of whether or not such vowels are reduced in a particular case can then be left till after the stress assignment.

The method outlined above follows linguistic processes more closely than that of some similar proposals (eg. Segre et al. [5]), in that it mirrors the fact that vowel reduction is directly predictable from the stress pattern, rather than indirectly from the orthographic representation. This faithfulness to linguistic principles also allows the method to serve as an explicitly-formulated means of testing a linguistic theory of word stress.

The prediction by rule of lexical stress placement for most words in English allows a speech synthesiser to handle an unrestricted range of input text. At the very least, it would be able to make a creditable attempt at pronouncing a word not encountered before - and even native speakers can often do no more than this.

Further developments and examples.

The program has only recently been implemented, and it is proposed to test it thoroughly with a large set of English words. These words will be drawn at random from a machine readable dictionary containing a phonemic transcription of each word with stress and hyphenation marked.

It is also hoped to conduct experiments with nonsense words, in order to compare the performance of the program with that of native English speakers. The results should indicate whether or not the principles followed by the program are in any way similar to the strategies used by native speakers in pronouncing previously-unknown words.

The morph-stripping algorithm has yet to be implemented, and much research is still needed in this area. Once such an algorithm is formulated, its output will form part of the input to the program described above.

Some examples are given below of input and output strings for the program in its present state: '<' is a strong prefix boundary, '> >' is a weak suffix boundary, and /[ɪ]/ is a centralised /ɪ/.

rɛ- < prɛ-zɛn-t > eɪ-ʃn̩ —> rɛ,-prə-zn̩-teɪ'-ʃn̩

hæn-dɪ-mæn —> hæ'n-d[ɪ]-mn̩

æ-lɒ- < mɔt —> æ'-lə-mɔt

mɒ-dɛ-r>eɪt [n] —> mɒ'-də-rət
mɒ-dɛr->>eɪt [v] —> mɒ'-də-reɪt

References

[1] Guierre. L. 1970 *Drills in English Stress-Patterns.* Paris: Armand Colin-Longman.

[2] Lawrence. S.G.C., Williams. B.J., & Kaye, G. (1984) The automated phonetic transcription of English text. (In this volume).

[3] Liberman. M. & Prince. A. (1977) On Stress and Linguistic Rhythm. *Linguistic Inquiry.* 8: 249-336.

[4] O'Connor. J.D. & Trim. J.L.M. (1953) Vowel. Consonant and Syllable - A Phonological Definition. *Word.* 9: 103-122.

[5] Segre. A.M.. Sherwood. B.A.. & Dickerson. W.B. (1983) An expert system for the production of phoneme strings from unmarked English text using machine-induced rules. In *Proceedings of the First Conference of the European Chapter of the Association for Computational Linguistics.*

[6] Selkirk. E.O. (1978) On Prosodic Structure and its Relation to Syntactic Structure. Paper presented at the Sloan Foundation Workshop on the Mental Representation of Phonology, University of Massachusetts. Amherst. November 18-19. 1979. Distributed by the Indiana University Linguistics Club, Bloomington. Indiana: 1980.

[7] Selkirk. E.O. (1980) The role of prosodic categories in English word stress'. *Linguistic Inquiry.* 11: 563-605.

[8] Selkirk. E.O. (1982) *The Syntax of Words.* Cambridge. Mass.. and London: M.I.T.

Acknowledgements

I should like to thank my colleagues in the Speech Research group for helpful discussions and in particular Mr Graham Lawrence. for his invaluable help in the development of this program. Dr Geoffrey Kaye (Manager of Speech Research) for his interest. and IBM for the post-doctoral fellowship.