Proceedings of The Institute of Acoustics

AUTOMATIC PHONETIC TRANSCRIPTION

G. Knowles

Dept. of Linguistics, University of Lancaster

This paper reports on work currently being undertaken at the University of Lancaster into linguistic problems of text-to-speech processing. The project is at a very early stage, and will be making use of earlier work at Lancaster on the grammatical analysis of written texts.

The kind of phonetic transcription envisaged is a detailed one, including stress and intonation, and features of connected speech. The aim is to find out how much is predictable from a written text, and how best to predict it. The phonemic make-up of words, for instance, is highly predictable for a given variety of English, and so are most allophones. On the other hand, intonation and pause are difficult to predict. Work so far has for this reason concentrated on transcribing words phonemically by interpreting regular English spelling conventions.

In the case of intonation, there are obvious limits to how much can be predicted. In view of the different ways in which a text can be interpreted, it will not be possible to predict in detail exactly how a human reader would read it aloud on a given occasion. Nor, in view of the 'attitudinal' meanings conveyed by intonation, will it be possible to make an exact reconstruction of a spoken original from a written version. The kind of intonation pattern that can be assigned belongs to that part of the system that might loosely be described as 'spoken punctuation'.

There are four major steps in making a detailed phonetic transcription of a text:

1. The grammatical analysis of the written version of the text.
2. The phonetic transcription of individual words.
3. Phrasing rules, grouping words into phrases.
4. Intonation rules.

1. GRAMMATICAL ANALYSIS

The grammatical analysis of the text involves both tagging and hypertagging. Whereas a tag indicates the part of speech of a word, a hypertag identifies a higher level constituent, e.g. prepositional phrase.

Tagging

The purpose of tagging is to obtain information necessary for hypertagging, and also for word transcription. Two tasks which can be done efficiently at this stage, are (1) identifying the "weight" of words, and (2) marking affixes.

Weight. The weight of a word determines its accentuation in context. All words are accented in isolation, but low-weighted words tend to lose their accent in context. WEIGHT can be thought of as an integer variable [0..4]:

4: nouns
3: lexical ("content") words other than nouns
2: normally accentable grammatical ("form") words, usually containing more than one syllable, e.g. inside, although
1: words unaccented but unreduced, e.g. on, they, I

0: words reduced to weak-forms, e.g. at, him.

It will be necessary to associate with each word in the text a weight value. This is predictable in the first place from the tag, e.g. that has weight 2 as a demonstrative, and 0 as a conjunction. It can, however, be altered in the light of higher level analysis, e.g. items in lists are not reduced, so the minimum weight for list items is 2. At normally has weight 0, but in final position - e.g. in What are you looking at? - it has weight 1. To also normally has weight 0, but 1 in cases of ellipsis, e.g. in I want to.
The deaccentuation of lexical words may in some cases be handled by reducing their weight to 1 or 0, e.g. -man has weight 1 in gasman and weight 0 in milkman.

By no means all part-of-speech problems can be handled by this one variable. A conservative noun like deBATE patterns like a verb at word level, but is still a noun at phrase level; and similarly an advanced verb like to COmment patterns as a verb at phrase level and as a noun at word level.

Affix-marking. To predict the accentuation of a word, it is necessary to analyze its morphology. The method presently used to pass on information about the position and type of affix boundaries is to insert a marker in the word:
(1) Germanic suffixes: these are 'neutral', and have no effect on accentuation, and are left out of account. All such suffixes need to be marked. The marker presently used is a dot, e.g. king.dom, hope.less.ness.
(2) Classical suffixes: just the last of these is marked. A heavy suffix (marked "+") is distinguished from a light suffix (marked "="), e.g. person+al, contro+versy, Americ=a, potat=o.

Although only the last Classical suffix needs to be marked, it is in fact necessary to identify all the Classical suffixes in a word. This is because a vowel letter immediately before a Classical suffix belongs (initially, at least) to a different syllable, so that the following syllable begins with a 'zero consonant', e.g. malari=_a, re_alit+y. The zero consonant is here marked with the underline character.
In addition, the letter <t> before the combination <i_> patterns not as a <t> but as a <c>, e.g. martial can be represented marci_al. A particularly important set of words involved here are those ending -ATION, e.g. relation (relaci_on).

Word transcription rules are marginally simplified if the plural/genitive <s> is changed to <z>, e.g. eggs becomes egg.z and cats cat.z.

Hypertagging
The value of hypertags is that they can be used
(1) to mark the beginnings and ends of structures relevant for intonation, and
(2) to mark off individual items in those structures.
The nature of these intonation structures is currently being investigated.

2. WORD TRANSCRIPTION

The rules to transcribe words phonemically operate on spellings as corrected in the course of morphological analysis. The program is divided into three main sections:
(1) graphemic analysis
(2) syllable division
(3) accentuation.

Graphemic analysis

AUTOMATIC PHONETIC TRANSCRIPTION

Although phonetic transcription might seem to involve essentially mapping strings of letters on to strings of phonemes, this is in fact a relatively minor section, dealing mainly with digraphs - e.g. sh, th, ch - and a few other oddities. This is because spellings contain several kinds of information which need to be retained at this stage: (1) about syllables, (2) about accentuation, and (3) sociolinguistic information. For instance, the double consonant in letter indicates that the preceding syllable ends with the vowel, and is accented. 'Hard <c>' is rewritten <k>, but 'soft <c>' remains as it responds differently from <s> to the effects of accentuation: compare face with /s/ and phase with /z/ following accented /ei/. And although words like sure, shore and Shaw may be pronounced alike in RP - and the spelling differences may seem to be irrelevant - the corresponding forms of other varieties of English can in fact be predicted from the spellings.

Syllable division

Syllable position is essentially a function of manner of articulation: the segment at the syllable boundary has a closer stricture than the segments on either side, and the syllabic is the segment with the greatest degree of opening. To deal with some phonotactic gaps, such as /tl/ or /pw/, provisional syllables have to be tested for legality, and the boundary shifted if necessary.
At this stage, "magic <e>" is still treated as a vowel: a word like name is consequently treated as a disyllable.

Accentuation

The accentuation or "stress pattern" of a word is determined in two stages. To begin with, all syllables are assumed to be accented, so that all accentuation rules involve the suppression of accent. Whether a syllable retains its accent or not depends on its weight:
(i) a LIGHT SYLLABLE ends with a short vowel, and
(ii) a HEAVY SYLLABLE either contains a long vowel, or ends with a consonant.
Light syllables are deaccented unless they are immediately followed by another light syllable. Syllables deaccented by this rule are also subject to vowel reduction rules, e.g. cinema, aroma. Light syllables which retain their accent may be subject to vowel lengthening rules, depending on the number of syllables following e.g. divine with /ai/ (before "magic <e>") but divinity with /i/. Subsequent rules deaccent syllables to the left of an accented syllable, and in final position. "Magic <e>" is also removed at this stage.

If word representations output by the accentuation rules were to be given a direct phonetic interpretation, they would be approximately half a millenium out of date. Accented vowels need to go through a set of rules which recapitulate sound changes of the last 500 years. Depending on which rules operate, and which are blocked, different varieties of English can be generated. Most of these rules involve vowels. Here are some typical examples:

Long vowels. Re-writing <i_e> as /ai/, <ee> as /i:/, or <a_e> as /ei/ recapitulates the Great Vowel Shift. Varieties of English differ in the subsequent development of long vowels before /r/, in words like sure or share. For Scottish English, most of the rules for vowels before /r/ are blocked.

Short vowels. In most words <u> is re-written /ʌ/, e.g. but, mud, but this rule is blocked for Northern English. After /w/, <a> is usually re-written /o/, e.g. was /woz/.

Lengthenings and shortenings. Before /k/, <oo> is usually shortened to /u/, e.g.

hook, but this rule is blocked for North-Western English. For Southern English, <a> is lengthened before /s/ or /f/ in the same syllable, e.g. mast, laugh.

Examples of developments in consonants include the simplification of <wh> (=/hw/) to /w/, e.g. <when> /wen/, and the loss of /g/ after the velar nasal in a word like sing. This latter rule is blocked for North-Western English.

Irregular spellings

Since the attempt is being made to interpret regular spelling conventions, the problem of irregular spellings has been deliberately postponed. As the rules are improved, some odd-looking spellings - e.g. biscuit, disguise - turn out to be regular.

There will inevitably remain a number of genuine irregularities. These can be amended at the point where they are irregular, so that they pass correctly through subsequent rules.

Where an anomalous spelling has arisen through medieval scribal practice, this can be amended at the stage of graphemic analysis. In words like son, woman, come, where <u> has been changed to <o> in the environment of a letter with several upright strokes, <u> can be restored. But many irregularities belong to the historical section, words having escaped the influence of sound changes, e.g. spook is an exception to the rule shortening <oo> before /k/. Similarly, great and steak are exceptions to the rule whereby <ea> comes to be pronounced like <ee>.

3. PHRASE RULES

Phrase rules deal with the segmental modifications of words when they are put together in phrases:
(1) assimilation,
(2) elision,
(3) hiatus, including linking and intrusive /r/.

These rules are fully discussed in standard textbooks such as Gimson [1], and need not be discussed in detail here.

4. INTONATION RULES

There are at least three types of rule for 'punctuating' by intonation:
(1) Rules for assigning onset and nucleus to the phrase,
(2) rules for combining phrases to form higher level structures, and
(3) 'lightening' rules.

Onset and Nucleus

Intonation rules proper start with the phrase:
(i) the last accented syllable is identified and marked as the NUCLEUS with the backslash "\".
(ii) The first accented syllable is identified and marked as the ONSET with the up-arrow "↑". (Onset and nucleus fall on the same syllable if there is only one accented syllable in the phrase.)
(iii) The accent on any syllables between onset and nucleus is suppressed.

Combination Rules

Rules combine phrases to form tone-groups, and tone-groups to form higher level intonation structures. These rules have yet to be formulated in detail.

Lightening rules

AUTOMATIC PHONETIC TRANSCRIPTION

The output of these combination rules is likely to be a very 'heavy' intonation, akin to a very heavy punctuation in a written text. For instance, fish and chips is indeed a list, but it is not usually necessary to enumerate its elements carefully!
Some lightening rules are automatic, and in some cases the accents of two tone-groups with one accent each can be collapsed into one tone-group as onset and nucleus. In other cases, set collocations and expressions (such as fish and chips) may have lighter intonation than newly coined ones.

REFERENCE

[1] A.C. Gimson, An Introduction to the Pronunciation of English. London: Edward Arnold. (3rd edition, 1980)

1.10.84