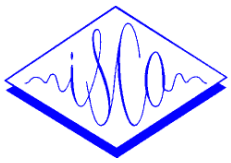

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

ACCENT IDENTIFICATION WITH A VIEW TO ASSISTING RECOGNITION

(Work in Progress)

Chris Cleirigh & Julie Vonwiller

Speech Technology Research Group, Dept. of Electrical Engineering,
The University of Sydney, NSW, 2006, Australia
Email: julie@speech.su.oz.au
cleirig@speech.su.oz.au

ABSTRACT

This paper reports on work in progress to identify accented Australian English varieties to facilitate identification. Accented speech sounds different because the choosing and chaining of segments in production of speech is influenced by a speaker's linguistic background. The paper describes the development of a descriptive model of the phonology of Australian English from the theoretical standpoint of Systemic Phonology. The schema proposes morphemes composed of an obligatory core syllable element and optional peripheral elements. Core syllables are unmarked for grammatical distribution, but peripheral elements are restricted in their distribution to morpheme boundaries. We propose that this gives us a more powerful tool for description and hence identification of the accents observed.We hypothesise that using this approach has several benefits for recognition systems.

I. INTRODUCTION

In Australia, approximately 25% of the population were born in another country, often speaking a language other than English as their mother tongue. Most of these people arrived in Australia after puberty and generally speak English with an accent. We are currently involved in collecting a national database of spoken Australian English (AuE) and would like to be able to more accurately describe what is meant by the appellation 'accented'. Further, our research group is interested in developing automatic recognition of accented Australian English. The results of pilot studies in this area undertaken by our group have been reported in Dowling et al. (1992), Blackburn et al. (1993) and Lewis (1993). For this work our aim is twofold- to provide a definition of accent and accent types and to build a model of accent components to facilitate machine recognition of accented Australian English. In this paper we report on the first phase of the project where we establish the frame of reference and the descriptors for the analysis. Our approach focuses on the syllable, and the theoretical stance will be that of Systemic Phonology.

Systemic Theory provides explicit relations between the levels of linguistic analysis -- semantics, lexicogrammar and phonology, and it is the contact of phonology with lexicogrammar -- specifically syllables with morphemes -- which is the area of interest to be studied here. The theory looks at the changing contexts of segmental selection during the speech process, and makes it possible to examine phonological patterning as constrained by lexicogrammatical context.

Our work reported here does not include a discussion of the role of stress which will be included in a later phase. We will look first at the syllable structure of English from the Systemic perspective, then consider this as it affects a non-native speaker of English using South Vietnamese accented English as a example.

II. ENGLISH SYLLABLE STRUCTURE

The types of syllable structures that express morphemes vary across languages. For English there are a number of ways in which syllables and morphemes relate. For example,

word	'paper'	'banned'	'statement'
morpheme	paper (1)	bann-ed(2)	state-ment (2)
syllable	pa-per (2)	banned (1)	state-ment (2)

English syllables have the structure Onset^Rhyme, where the Onset may be realised by a single consonant or expanded as a double-consonant cluster $C_1(C_2)$, and the Rhyme may be realised by a vowel or expanded as a vowel-consonant cluster $(V(C_3)$. The internal structure of the syllable is captured by specific phonological patterning which is the result of two dimensions of the phonological process, the paradigm and the syntagm. *The paradigm refers to the system of choice of phonemes/ segments for inclusion in the syllable, and the syntagm refers to the system of chaining of these phonemes/ segments within the syllable.* These will be examined from a perspective of phonological-lexicogrammatical interaction. The table below sets out the relationships. The theory has been developed by Cleirigh (forthcoming) for Irish and is being extended to

English in this project.

2.1 Morpheme-syllable Interaction

The stratal relation between lexicogrammar and phonology in Systemic theory is as follows. Lexicogrammar is realised by phonology and provides a higher level context for phonological selections.

Considering, then, morpheme-syllable interaction, morphemes are usually realised by syllables (1 or more), and provide higher level contexts for syllable selections.

Syllables can be said to occur in two principle lexicogrammatical contexts:

1) morpheme-initial, as **'le'** in the English disyllabic morpheme **'le**vel', and

2) morpheme-final, as **'vel'** in 'le**vel**'.

English syllable structure potentially reflects these two lexicogrammatical contexts, and there are, therefore, three types of syllable structures, namely:

1) syllables that do NOT reflect any lexicogrammatical context: the so-called 'core syllable',

2) syllables that reflect morpheme-initiality: through an additional peripheral segment, so-called 'proclitic /s/', and

3) syllables that reflect morpheme-finality: through an additional peripheral segment or segments, so-called 'enclitic consonants'.

Table 1: types of syllable structure in English

lexicogrammatical context	
unspecified	example
'core' $(C_1(C_2))^V(C_3)$	/trey/ 'tray'
morpheme-initial	example
/s/^ 'core' $s^(C_1(C_2))^V(C_3)$	/strey/ 'stray'
morpheme-final	
'core' ^ C(C) $(C_1(C_2))$ ^ $V(C_3)^C_4(C_5)$	/treyt/ 'trait' /treyps/ 'traipse'

The above table sets out the three possible syllable structures: Note that the numbering of consonants indicates that the range of paradigmatic choice varies for each. More specifically:

C_1: any consonant other than /N[1]/.

C_2: one of /y w r l/, but not all may combine with all C_1s.

C_3: one of /y w r l n s/ where /y w r/ result in long vowels and diphthongs, and /n/ is [n] by default but assimilates place of articulation to a following stop consonant.

C_4: any consonant other than /h/, but not all may combine with all C_3s

C_5: usually one of /t s/

After simple Rhymes, that is after the core of the syllable, any consonant -- except /h/ -- can occur as an 'enclitic'. It would seem that the maximum number of (morpheme- final) enclitics for a single syllable is two or three, depending on whether we treat such clusters as / ks/ in 'text' and /ps/ in 'glimpse' or 'lapse' as a single affricate or as two consonants. After expanded (complex) Rhymes the choice and number of morpheme-final selections is more constrained.

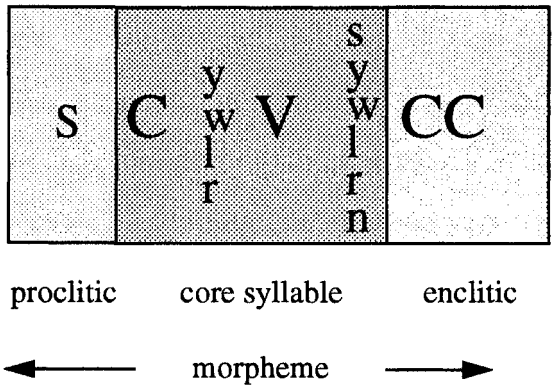

Figure 1. schema for English phoneme selection

There is a distinction between syllable structures (phoneme sequences) that have free occurrence from those that are restricted to specific lexicogrammatical contexts. For example, in English, morphemes can be expressed by one or more syllables, but certain phonemes in particular syllable positions indicate morpheme boundaries. For example, 't' can only close syllables that are morpheme-final. By referring to the diagram above, we can see that 't' is not part of the extended Rhyme of the core syllable, so segment 't' closing a syllable must be an enclitic consonant which only occur morpheme-finally. Therefore in a word

1. SAM symbols are used here and in the examples on page 4.

realised by two syllables, for example 'catnap', in which the first syllable ends in 't', the first syllable (cat) must also be a morpheme. Thus this approach provides a method of predicting morpheme boundaries in certain cases and thus the morphological structure of words.

All the possible core syllables were calculated and there were found to be of the order of 750 of these freely occurring syllables, which is a considerable reduction on the number of syllables unusually listed (2-3000).

At the level of syllable morpheme interaction, two types of phenomena can be identified as occurring 'within the morpheme' but 'around the syllable'. These correspond to what Firth (1948) termed 'extension prosodies' and demarcation prosodies'. The discussion to date has focused on demarcation prosodies, a typical phonological strategy for signalling the domain of the morpheme for English. Extension prosody can be seen in vowel harmony in languages like Turkish, and in tone in languages like South Vietnamese. Languages use these strategies to different extents. Extension and demarcation prosodies can be thought of as having lost their association with the domain of the syllable and having become associated with a lexicogrammatical domain of the morpheme. They express the cohesion of the lexicogrammatical unit (morpheme) by indicating its domain through extension, which integrates the unit, or through demarcation of its boundaries.

III. ACCENTED ENGLISH

The above pattern represents the core and periphery structure of the basic syllable for English, and demonstrates that English makes use of demarcation prosody in establishing the integrity of the morpheme. One could predict that these are language specific, and we postulate that the degree of accent will represent the degree of mismatch of the paradigm and syntagm of L1 when speaking L2. In languages where the segment inventories are quite different there will be difficulties with the paradigmatic choices. In languages where the syllable-morpheme interaction is governed by chaining choices in L1 that are different to that of L2, say English, then there will be difficulties with the syntagmatic structure. Where both the paradigm and the syntagm differ, then the degree of difficulty will be greatest. We postulate 3 types of accents corresponding to these levels of mismatch and predict that the paradigmatic type will cause the least interference to the human listener. Type 1 is a paradigmatic accent; type 2 is a syntagmatic accent; and type 3 is a combined paradigmatic/syntagmatic accent.

For the recognition task, we now have the following advantages, (a) a reduced set of syllables, (b) the core syllable is primarily affected by paradigm substitutions and is therefore the least affected by accent, (c) paradigmatic mismatches are generally dealt with by substitution of an L1 approximation not omission. This suggests that the core syllable will be most likely identified and should be the focus of recognition of accented Australian English. As a preliminary test our hypothesis, we examined the speech of a South Vietnamese-born Australian speaking the 200 Australianised SCRIBE sentences. These were a section of the collection of the Australian National Database of Spoken Language (Millar et al 1994)

3.1 South Vietnamese

Vietnamese uses tone as an extension prosody, and, perhaps consequentially, makes less use of demarcation than English. .The maximal structure of the Vietnamese syllable, Onset^Rhyme where the Onset may be realised by a single consonant or expanded as a double-consonant cluster $C_1(C_2)$, and the Rhyme may be realised by a vowel or expanded as a vowel-consonant cluster ($V(C_3)$ can be represented in segmental terms[1] as:

$(C_1(C_2))V(C_3)$ where () indicates optionality.

This structure matches the 'unspecified' type of syllable structure of English in Table 1. Each consonant position is differentially constrained for choice.

C_1 can be realised by any consonant.

C_2 may only be realised by /w/, and not all C_1s can be so modified. Those C_1s that can't be modified by /w/ are all labials, liquids and dental /n/.

V can be realised by any vowel.

C_3 position can be realised by

(1) the (unreleased) voiceless stops /p t k/, with /t/ and /k/ in complementary distribution;

(2) the nasals /m n N/, with /n/ and /N/ in complementary distribution.

3.2 South Vietnamese Accented English

The accent of the South Vietnamese will be considered from the point of view of the syllable structures of English in Table 1. The phoneme inventories for AuE and South Vietnamese are similar, though Vietnamese has a trill, a retroflex series, an extra nasal, and velar fricatives and lacks dental fricatives. For the test subject the 'unspecified syllable structure' was not a major problem. The paradigmatic choice that presented the

1. This analysis is based on a segmental description, but an alternative (prosodic) approach suggests that Vietnamese makes use of posture transitions that complicate an underlyingly simple segmental inventory.

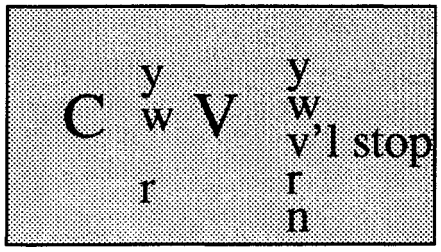

Figure 2. schema for Sth. Viet. phoneme selection

most difficulty was /l/. The most obvious difficulty was with the phoneme /l/ in C_2 position and in C_3 position. Both were positions in which the sound did not occur in South Vietnamese. e.g. 'clouds' /laus/; 'chill' /tSau/.

With the 'morpheme-initial structure' he fluctuated between achieving a proclitic /s/ and omitting it. This was taken as evidence that L1 still interfers with L2 even after several years of speaking L2. Such variance in the learning process is a common feature and must be taken into account in this kind of work. Examples:-

'staring' /tEriN/
'scar' /ke:/
but
'spring' /sprIn/

Most of the subjects difficulties were with the 'morpheme-final structures' and these problems were consistent throughout his performance. Here the speaker deleted C_3 following a long vowel or diphthong, that is, the consonant was outside of the core and the subject's language makes no use of enclitics. Seee the examples below.

bedside' /bEsai/
'ancient' /Ansen/
'statement' /steim@n/
'present' /preis@n/
'ointment' /oimEnt/
'like' /lai/
'since' /sIn/

Interestingly, when the word contained a segment representing an additional morpheme such as a plural or possessive 's', the 's' would be retained in preference to segments in the preceding morpheme. For example,

'Yates' /jeis/
'Brown's' /braus/

IV. 'SUMMARY

This is a brief introduction to the study of Systemic Phonology and its application to the study of accent. The procedure is still under development and the application to other types of accent will be undertaken in the near future. We feel that this approach has great potential to provide a tool for systematic analysis of accented speech. Also, we feel that there is evidence that this approach may be useful in classifying degrees of accent.

REFERENCES

[1] Bazell CE, Catford JC, Halliday MAK & Robins RH (1966) In Memory Of JR Firth London: Longmans

[2] Blackburn C., Vonwiller J. & King R. (1993) "Automatic Accent CLassification using Artificial Neural Networks" Eurospeech'93. 1241-1244.

[3] Clark J & Yallop C (1990) An Introduction To Phonetics And Phonology Oxford: Basil Blackwell

[4] Cleirigh C. (forthcoming) "Phonic Texture: A Systemic Approach to the Syllable" Doctoral Thesis, Dept. Linguistics, The University of Sydney.

[5] Dowling A., King R. & Vonwiller J. (1992) An Artificial Neural Network for Automatic Spoken Language Recognition" SST 92 pp156-161.

[6] Firth JR (1948) Sounds And Prosodies in Palmer FR (ed) (1970) Prosodic Analysis London: OUP

[7] Halliday MAK (1985) An Introduction To Functional Grammar London: Edward Arnold

[8] Halliday MAK (1992) A Systemic-Prosodic Account Of Peking Syllable Finals in P Tench(ed) 1992 Studies In Systemic Phonology London: Pinter

[9] Henderson E (1966) Towards A Prosodic Statement Of Vietnamese Syllable Structure in Bazell et al (1966) In Memory Of JR Firth London: Longmans.

[10] Lewis W.J. (1993) "A Preliminary Study of Western Slavic-Accented English" Unpublished Honours Thesis, Linguistics Dept. The University of Sydney.

[11] Millar,J.B., Vonwiller,J.P., Harrington,J.M., Dermody,P.J (1994) The Australian National Database of Spoken Language, Proc ICASSP-94 Vol 2: pp 67-100

[12] Nguyen Dang Liem (1970) " A Contrastive Phonological Analysis of English & Vietnamese" Pacific Linguistics: Series C. No. 8

[13] Palmer FR (ed) (1970) Prosodic Analysis London: OUP

[14] Tench P (ed) 1992 Studies In Systemic Phonology London: Pinter.