NOVEL-WORD PRONUNCIATION: A CROSS-LANGUAGE STUDY

K.P.H. Sullivan

Department of Computer Science
University of Otago

ABSTRACT – In the case of a 'novel' word absent from a text-to-speech system's pronouncing dictionary the traditional systems invoke letter-to-phoneme rules to produce a pronunciation. A proposal in the psychological literature, however, is that human readers pronounce novel words not using explicit rules, but by *analogy* with letter/phoneme patterns for words they already know. A synthesis-by-analogy system is presented which is ,accordingly, also a model of novel-word pronunciation by humans. The computational methods of assessing the orthographic analogy module and the 'flexible' (context-independent) GPC rule module, a pre-requisite for phonological analogy, are presented. The resultant assessments across language, method of assessment, size and content of the lexical database are compared, before implications the future development of computer synthesis-by-analogy and for psychological models of oral reading are presented. The investigations into these modules produced useful results for both British English and German.

INTRODUCTION

Novel-word pronunciation has widely been used as a means of assessing and developing psychological models of oral reading (e.g. Glushko, 1979). The development of such models can be assisted by computational modelling of reading. Such modelling affords the possibility of controlling variables not easily controlled when experimenting with human subjects.

Sullivan and Damper (1990; 1992) have previously described a text-to-speech system based on the experimental work of Glushko (1979) and Brown and Besner (1987). Glushko posited that human beings pronounce words not contained in their personal lexicon ('novel-words') by analogy with the entries in their orthographic lexicon. Brown and Besner, on the other hand, believe their experimental results show that phonological similarity is the kernel of analogy. Sullivan and Damper's model employs analogy in both orthographic and phonemic domains together with a means of resolving conflicts between the two (see Figure 1).

Most other computer text-to-speech systems use a dictionary of pronunciations conjunction with a set of grapheme-to-phoneme conversion (GPC) rules employed for novel-words – those not in the dictionary

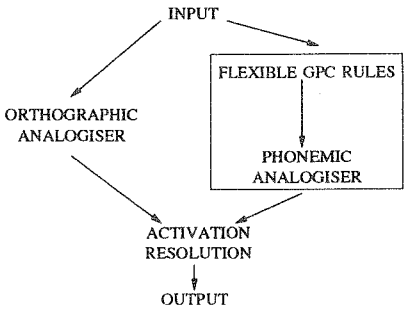

Figure 1. Schematic diagram of the synthesis-by-analogy model. The analogy process is applied in both orthographic and phonemic domains. The latter requires the use of 'flexible' GPC rules (see text).

of pronunciations. To avoid over-generality, leading to multiple candidate pronunciations, the rules are context-dependent (i.e. left and right contexts must be matched before a rule is applied). The rule-based approach to text-phonetics conversion has many problems; most importantly, it is unclear how to derive the *canonical* set of GPC rules on which the system is critically based.

This 'dual route' approach to text-to-speech conversion to a certain extent mirrors an early psychological model of reading aloud (e.g. Coltheart, 1978) in having both lexical and rule-based routes to pronunciation. Subsequently, Glushko (1979, 1981) made his revolutionary proposals about analogy. In spite of its intuitive appeal, however, synthesis-by-analogy remains a little explored topic for computational modelling or text-to-speech systems.

This paper explores the computational methods of assessing the orthographic analogy module and the 'flexible' GPC rule module of Sullivan and Damper's model. The 'flexible' GPC rule module which produces the set of plausible, candidate pronunciations phonemic analogy requires. (Here 'flexible' indicates that the rules are context-independent.) The resultant assessments are compared across language (German and English), method of assessment, size and content of the lexical database. Thereafter, implications for the future development of computer synthesis-by-analogy and for psychological models of oral reading are considered.

COMPUTATIONAL SYNTHESIS-BY-ANALOGY

The 'flexible' GPC and the orthographic analogy modules operate in essentially the same manner; they differ only in the variety of unit size under consideration. Both modules require a means of evaluation to enable selection from the competing generated pronunciations.

Preference Value

A given orthographic substring o can map to a number of possible phonemic substrings p. Let the number of such correspondence be C. The probability that orthographic substring o maps to phonemic substring p_j, given o, is estimated by the total number of $o \rightarrow p_i$ correspondences in the corpus, normalised by the total number of correspondences involving o. Thus:

$$P(o \rightarrow p_j | o) = \frac{\sum_{w=1}^{L} F_w N_w(o \rightarrow p_j)}{\sum_{w=1}^{L} \sum_{i=1}^{C} F_w N_w(o \rightarrow p_i)} \quad (1)$$

Here, $N_w(o \rightarrow p_j)$ is the number of $o \rightarrow p_j$ correspondences in word w, and F_w is the frequency of word w in the lexicon of size L. For the British English implementations we use the correspondences tabulated by Lawrence and Kaye (1986) and an alignment algorithm similar to theirs. For the German implementations we developed a set of correspondences using the same criteria as Lawrence and Kaye in the development of their correspondences for British English. The resultant probability-like values, which reflect the likelihood of that particular grapheme-to-phoneme conversion occurring, are referred to as the *preference value*.

Lawrence and Kaye (1986) list two set of statistics #text and #lex. The #text statistic gives the number of words in the Lancaster-Oslo/Bergen corpus (LOB) which contain a particular correspondence, whereas the #lex gives the number of times that a particular correspondence occurred in the Collins English Dictionary. These statistics allow a set of conditional mapping probabilities to be computed. For instance, *a posteriori* probabilities conditioned on the occurrence of orthographic substring o are:

$$P_{text}(o \rightarrow p_j | o) = \frac{\#text(o \rightarrow p_j)}{\sum_{i=1}^{C} \#text(o \rightarrow p_i)} \quad (2)$$

$$P_{lex}(o \rightarrow p_j|o) = \frac{\#lex(o \rightarrow p_j)}{\sum_{w=1}^{L} \#lex(o \rightarrow p_i)} \quad (3)$$

$$P_{prod}(o \rightarrow p_j|o) = \frac{\#text(o \rightarrow p_j)\#lex(o \rightarrow p_j)}{\sum_{w=1}^{C} \#text(o \rightarrow p_i) \sum_{i=1}^{L} \#lex(o \rightarrow p_i)} \quad (4)$$

These equations thus provide us with values which can be used as a basis on which to select candidate pronunciations. It should be noted, however, that those values based on the #text statistics ignore multiple occurrences of a correspondence in the same word and that those based on the #lex statistics assume that all words have the same frequency of occurrence.

Analogy Operation

To produce the set of candidate pronunciations for an input word all possible letter to phoneme conversions are considered. These are grouped according to the position of the initial letter within the input word. For the example pseudoword *pook* these are:

Group 1) *p, po, poo, pook;*
Group 2) *o, oo, ook;*
Group 3) *o, ok;*
Group 4) *k.*

In the case of the flexible GPC convertor, most of these orthographic substrings will not be members of the GPC rule-set – they did not along to a phoneme cluster during data alignment. Ordering by length in this way permits the remaining substrings within the group under consideration to be disregarded as soon as the substring under consideration does not invoke a conversion to a phonemic transcription. The process then moves to the next group. Considering the 'flexible' GPC module, after successful conversion of *p* in Group 1, the attempted conversion of *po* will fail (there is no GPC rule for *po* and therefore none for any of the remaining letter clusters of group 1). The next substring considered will be the first member *(o)* of Group 2. If a rule did exist for the cluster *poo, po* would have transcribed to *NIL*, with no attached preference value.

Whenever a graphemic substring is successfully converted into phonemes, these phonemes are entered into a pronunciation lattice for the input word. The lattice contains phonemic outputs and their preference values, along with information indicating which grapheme sub-string produced each lattice entry. Once all possible substrings of the input word have been processed in this way, candidate pronunciations can be generated. The enumeration of the possible paths through the lattice from start to end is conveniently done by using path algebra (Carré, 1979).

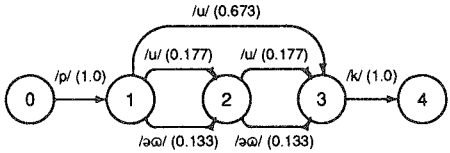

Figure 2. Pronunciation lattice produced by phonemic analogy for the pseudoword *pook*. The nodes are the junctures between letters and the arcs therefore represent orthographic-to-phonemic mappings. Arcs are labelled with corresponding phonemic substrings and, in brackets, mapping probabilities (see text).

Figure 2 shows a simplified illustration of the possible enumerated paths through the pronunciation lattice for the input *pook*, given the GPC rules and their mapping probabilities (shown in brackets):

$p \rightarrow$ /p/ (1.0);
$o \rightarrow$ /u/ (0.177),
$o \rightarrow$ /əʊ/ (0.133);
$oo \rightarrow$ /u/ (0.673);
$k \rightarrow$ /k/ (1.0).

At each node the lattice must be traversed by an arc labelled with both a phonemic transcription and a preference value. Since a large number of candidate pronunciations are generated, they are narrowed down to the best twenty prior to their consideration by the analogy stage. This method is also used by orthographic analogy to select the best pronunciation.

Hence, for the simple illustrative example, based on *pook*, the candidate pronunciations are /puuk/, /puəʊk/, /pəʊuk/, /pəʊəʊk/ and /puk/. These candidates are quantitatively evaluated using a (probability-like) value called the *confidence rating*. Many way of obtaining a confidence rating have been investigated (Sullivan, 1992). The best way of doing this was found to be to simply take the product of the preference values for the invoked correspondences. This gives the confidence values of 0.0313, 0.0235, 0.0235, 0.0177 and 0.673, respectively, for the above candidate pronunciations. These values are then used to rank the candidates.

CANDIDATE IMPLEMENTATIONS

To provide a basis for module and candidate implementation comparison pseudowords were presented to 20 native speakers for British English and 10 native speakers of German. The English pseudoword set consisted of 136 words created by Glushko (1979) by changing the initial consonant or consonant cluster of a monosyllabic word. The 100 German pseudowords were created by the same technique. The speakers' pronunciations were transcribed into IPA. In scoring the system's outputs a 'correct' pronunciation was one which was pronounced by any of the human subjects. A pseudoword has no correct pronunciation; people will pronounce such words differently. Therefore, the most frequent pronunciations cannot become the target pronunciations; this would deny the validity of some people's pronunciations.

British English

Initial investigation into the treatment of word-initial and word-final graphemes (Sullivan and Damper, 1991; 1992) suggested that these graphemes were best treated as both equivalent and distinct. The modules considered were: module 1 treated word-initial and word-final graphemes as equivalent, e.g. for the lexical entry *green*, probability calculations are based on the alignment form $g \rightarrow$/g/,...,$n \rightarrow$ /n/. Module 2 considers initial and final grapheme to be distinct. That is, the alignment form is $\$g \rightarrow$ /g/,...,$n\$ \rightarrow$ /n/, where \$ is the word delimiter. Finally module 3 treated initial and final graphemes as both equivalent and distinct. The alignment form is $\$g \rightarrow$ /g/, $g \rightarrow$ /g/, ..., $n \rightarrow$ /n/,$n\$ \rightarrow$ /n/.

This initial work was based on the 800 words of Ogden's (1937) Basic English and the word frequency statistics from Kučera and Francis (1967). Two further databases were constructed to investigate the effect of varying the lexical database and to confirm the results from the examination of word-initial and word-final graphemes. One was based on the 800 most frequent words of the Kučera and Francis corpus (KF800) and the other on the 3926 words in the Oxford Advanced Learners' Dictionary (OALD) (1989) with a frequency of 7 or above in the Kučera and Francis corpus.

The treatment of word-initial and word-final graphemes as both equal and distinct did not result in the highest number of top-ranked pronunciations in the case of the OALD-based flexible-rule based module or the KF800-based orthographic analogy module. Nevertheless the overall best results, when using the $P(o \rightarrow p_j|o)$ values, are obtained by the 800 words for Basic English for the flexible GPC module

3, with 58.8% rank 1 correct pronunciations, and by the OALD for the orthographic analogy module 3 with 64.7% rank 1 correct pronunciations. It is not unexpected that the implementations based on the 800 most-frequent words of the Kučera and Francis corpus performed least well since this lexicon contains all high-frequency words which are likely to be less regular in their pronunciation, than the equally-sized 800 words of Basic English which contains a cross-section of content and function words. Equally it is not surprising that the best-performing orthographic analogy implementation was based on the Oxford dictionary. This was the largest database and therefore contained the best cross-section of analogy segments. It is surprising, however, that the flexible GPC modules based on the same dictionary performed as badly as they did.

The #text and #lex statistics provided by Lawrence and Kaye (1986) present a way of assessing the effects of a larger database on the flexible GPC module. The score of 72.1% top-ranked pronunciations for the *a priori* implementation using Equation 4 values is highly surprising not only because it outperforms the best flexible GPC module (which uses the most principled equation 1) by 17.7% percentage points, but also since the *a priori* implementation outperforms the *a posteriori* implementation.

This result could be due to Lawrence and Kaye's statistics capturing more of the regularities of English spelling-sound correspondences, as they are based on a much larger lexicon. However, the *a priori* type 1 module implementation using equation 4 values based on the 800 words of Basic English and the statistics of Kučera and Francis was the top-performing flexible GPC modules with 78.8% rank 1 pronunciations. Orthographic analogy also performed better using equation 4 values. A type 3 *a posteriori* implementation based on the 800 words of basic English and Kučera and Francis resulted in the best orthographic analogy performance of 70.6% rank 1 pronunciations.

German

Similar modules were constructed for a German synthesis-by-analogy investigation. The database consisted of the 800 most-frequent words according to Meier's (1967) *Deutsche Sprachstatistik*. The top-ranking flexible GPC modules was a type 2 implementation based on equation 1 values with 49% top-ranked pronunciations and the best orthographic analogy module was a type 3 implementation also based on equation 1 with 82% rank 1 pronunciations.

CONCLUSIONS

The investigations into these modules produced useful results for both British English and German. These successful results mask the uncertainties underlying the operation of these modules and the contradictory means of how best to produce effective synthesis-by-analogy. Further work is required into the information required to calculate the preference value and to resolve the question as to whether the information required is language dependent – our English and German results point to a substantial difference. Equally investigation into more and larger lexica may positively indicate how best to treat word-initial and word-final graphemes for flexible GPC and orthographic analogy modules – they may be different.

The British English 'flexible' GPC implementations all produced more top-ranked pronunciations than their respective orthographic analogy modules. This not only raises questions about computational orthographic analogy but also questions the conclusion drawn by experimental psychologists. The best German orthographic analogy implementations produced a score 30 percentage points greater than the top-scoring flexible GPC implementation. This change in performance is possibly connected with the problem; the relationship between letter and sound is more direct in German than in British English. This warrants further investigation.

It has been shown that the performance of a database is not primarily dependent upon its size, but rather its contents, which can be easily controlled in a computational model. Therefore, investigation into whether different lexica (sub-sets of the mental lexicon) are consulted depending upon the situation in which the novel-word is encountered is planned. For example, if a novel-word is perceived to be French only the French sub-set lexicon may be consulted to generate the pronunciation for that

particular novel-word. If this research shows that sub-set lexica are consulted in such situations, then it is probable that general isolated pseudowords are pronounced through the use of a sub-set of general words. This would explain the better performance of the Basic English module over that based on the Oxford Advanced Learner's Dictionary.

REFERENCES

Brown, P. & Besner, D. (1987) "The assembly of phonology in oral reading: A new model," in Coltheart, M. (ed.), *Attention and performance XII: The psychology of reading*, pp. 471–489, (Lawrence Erlbaum: London).

Carré, B.A. (1979) *Graphs and networks*, (Oxford University Press: Oxford, UK).

Coltheart, M. (1978) "Lexical access in simple reading tasks," in Underwood, G. (ed.), *Strategies of information processing*, pp. 151–216, (Academic Press: London).

Glushko, R.J. (1979) "The organization and activation of orthographic knowledge in reading aloud," Journal of Experimental Psychology: Human Perception and Performance, 5, 674–691.

Glushko, R.J. (1981) "Principles for pronouncing print: The psychology of phonography," in Lesgold, A.M. & Perfetti, C.A. (eds.), *Interactive processes in reading*, pp. 61–84, (Lawrence Erlbaum: Hillsdale, NJ).

Kučera, H. & Francis, W.N. (1967) *Computational analysis of present-day American English*, (Brown University Press: Providence, RI).

Lawrence, S.G.C. & Kaye, G. (1986) "Alignment of phonemes with their corresponding orthography," Computer Speech and Language, 1, 153–165.

Meier, H. (1967) *Deutsche Sprachstatistik*, (Georg Olms: Hildesheim).

Ogden, C.K. (1937). *Basic English*, (Kegan Paul: London).

Oxford University Press (1989) *Oxford Advanced Learner's Dictionary of Current English, 3rd Edition, Electronic Handbook*, (Oxford University Press: Oxford, UK).

Sullivan, K.P.H. (1992) *Synthesis-by-Analogy: A Psychologically-Motivated Approach to Text-to-Speech Conversion*, PhD Thesis, University of Southampton, Southampton, UK.

Sullivan, K.P.H. & Damper, R.I. (1990) "A psychologically-governed approach to novel word pronunciation within a text-to-speech system," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '90, Albuquerque, NM, Vol. 1, 341–344.

Sullivan, K.P.H. & Damper, R.I. (1991) "Speech synthesis by analogy: Recent advances and results," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '91, Toronto, Canada, Vol. 2, 761–764.

Sullivan, K.P.H. & Damper, R.I. (1992) "Novel-word pronunciation within a text-to-speech system," in Bailly, G & Benoît ., (eds) *Talking Machines: Theories, Models and Applications* pp. 183–195, (Elsevier: Amsterdam).