# A RULE LEARNING SYSTEM AND ITS APPLICATION TO MACHINE SPEECH

S. OAKEY AND R. C. CAWTHORN

DEPT. OF COMPUTER SCIENCE, TEESSIDE POLYTECHNIC

## Introduction

A number of researchers have been interested in the use of a set of pronunciation rules for speech generation, and two such rule sets (1) and (2) have been reported. Both rule sets were produced manually, essentially by checking the rule set against some dictionary, noting the incorrect pronunciations, altering the rules, and re-iterating. The research reported here has automated this process.

## Rule production

A word is read in with an associated pronunciation. The current rule set is applied to the word to produce an hypothesised pronunciation, and this is compared with the input pronunciation. If these are the same, no action is taken. If they are different, new rules are produced to augment the rule set.

For example, the word PALE would (initially) be pronounced "p ae l e" (in International Phonetic Alphabet characters). This is compared with the input phonetic spelling "p eI l" and special rules applicable only to that word are produced:

1. 'A' preceded by an initial 'P' and terminated by 'LE' is pronounced as 'eI' or, in the computer's rules format:
   (A (SPACE P) (L E SPACE) (eI))

and a second rule produced is

2. (E (SPACE P A L) (SPACE) NIL)

to denote that the final 'E' is silent.

## Rule induction

When specific rules like 1 and 2 above are produced, the system scans its current rule set for rules that are similar to the rule(s) produced. For example, if it already had the rule

3. (A (SPACE B) (N E SPACE) (eI))

it would compare rules 1 and 3 and produce the more-general rule

4. (A (SPACE CONS) (CONS E SPACE) (eI))

which also correctly translates words like SALE, SAFE, and RAKE.

This is the simplest form of generalisation, and the system rarely finds rules conveniently of the same length. Indeed, it needs to combine rules of different lengths in order to generalise towards the typical "long vowel" rule, for example. The inductive process in fact relies on specific rules being of

different lengths to produce useful rules of the type
5. (E (VOWEL CONS) (SPACE) NIL)

The process is to allow two or more rules to be combined to produce a general rule as long as they do not differ too much, and of course they must produce the same phoneme. The left and right contexts of the rules are compared, and elements that are equivalent are retained as the context of the general rule. For example, "EKE" might produce the rule
6. (E (SPACE E K) (SPACE) NIL)
and this, when combined with rule 2 above, generalises to give the silent 'E' rule, rule 5.

However, if no restrictions are placed on this process, over-general rules are produced. One of the controls placed on the process is the number of "discarded" terms from a rule's contexts. In the production of rule 5, rule 6 discarded one term (the left-hand SPACE), and rule 2 discarded two terms (the SPACE and the P). For the small dictionary (3), a maximum of two (as in this example) has kept the generalisation within reasonable bounds, though this maximum is dynamically altered by the system for its correction processes.

## Rule correction

A garbage collector was introduced into the system to look for redundant rules. Once the dictionary had been processed, this was used to re-read the words and check that all rules in the set were now being used. However, it reported that words that were correctly translated in the original run were later being translated incorrectly. The system was then modified to make it check its own rule set and, where necessary, amend it.

The process used begins by re-reading the words and their pronunciation. The rule set is used to hypothesise the pronunciation, expecting to agree with the input pronunciation. Where it does not agree, the system produces a special rule to correct the set. Knowing that a rule to handle the exception must have existed in the set, it is searched for and action is taken to bring it back into use, if that is possible. The usual reason for the error in translation is that an over-general rule has been produced, though occasionally it is because the rules are incorrectly ordered.

Each general rule includes a complete history of its formation, with the component rules appearing as a tree structure below it. If, therefore, a general rule is identified as being incorrect, it is "broken" into its components, each of which is re-entered into the rule set. These are still eligible for generalisation, and if no further action was taken would recombine to form the same over-general rule. The system therefore increases the "similarity" requirement before generalisation is allowed to take place, and this leads to over-general rules being reduced back to much better, less general rules.

This process was used on the small development dictionary. The first re-reading of the dictionary produced a number of changes to the rule set, but a second re-reading produced no changes at all. The system had stabilised, producing a rule set that translated as required all of the words (about 450) that were in the dictionary. The rule set produced was independent of word order in all but minor details, and was very similar to the rule sets quoted in (1) and (2), although the size of the dictionary has not allowed the rules to develop to the generality of the latter.

## Applications

The rule induction system is a very general tool. It essentially produces rules governing the translation of one input string into another, using examples of the translation to deduce the rules. One extension to the current study is to investigate the existence of a suitable rule set for English, and to investigate the proportion of words that can be "correctly" pronounced by any general rule set. The system is particularly useful in that the user can define a universe of discourse and produce a perfect rule set for that universe - though the system may show that it is cheaper to use direct dictionary look-up, with the word set taking less space than the rule set.

The application of the system in speech extends well beyond this, however. One application would be to reverse the input and output of the present system ie to investigate the existence of spelling rules. Obviously there would be spelling ambiguities, and since the system does not use sentence context they could not be resolved by it as it stands.

## Conclusion

The system is now undergoing tests with a full-size dictionary (4), consisting of some 20,000 words. This dictionary does confound two separate problems: the translation of free-standing unstressed words; and the effects of duration and stress. For example:

7. "EXAMPLE" is pronounced "I g z a m p l"
   rather than "e k s a m p e l"

The system, of course, produces a rule set that reflects this confounding, and this makes it difficult to compare its rule set with previous rule sets (which were, anyway, formed from a more-pragmatic approach). It may well be that the use of unstressed free-standing word pronunciation would give a better insight into an acceptable rulr set for English, and the system could be used to investigate th 'lower-level" pronunciation changes that occur in free-flowing speech as a separate study.

The system would also be more practicable were it modified to allow the hypothesis checking to be successful if one of a number of pronunciations was produced, rather than allowing only one dictionary pronunciation to be acceptable. This would certainly reduce the size of the rule set, and would be closer to the pragmatic approach of (2).

A particularly useful feature of the system is that it appears to be language-independent. The only constructs it uses are the consonant and vowel (although other workers have made use of other constructs in addition to these), with no known language-dependent feature built into it.

The size of the development dictionary, and its nature, has allowed a few problems to remain hidden, and the processing of the much larger dictionary has brought them to light, but they are practical rather than theoretical problems.

The system does successfully learn rules governing the pronunciation of words from examoles, and this has practical applications particularly where a predetermined set of words is to be used for speech synthesis. It also allows studies to be made of the economics of rule sets versus dictionary look-up.

References

1. W. A. AINSWORTH 1974 IEEE Trans. Audio Electocoust. Vol. AU 21, 288-290
A system for converting English text into speech.

2. S. ELOVITZ, R. JOHNSON, A. McHUGH and J. E. SHORE 1976 IEEE Trans. Vol. ASSP-24, 446-459.
Letter to sound rules for automatic translation of English text to phonetics.

3. W. MURRAY.
The Ladybird key words reading scheme.

4. J. W. LEWIS Oxford University Press.
A concise pronouncing dictionary of British and American English.

5. A. WIJK 1966 Oxford University Press.
Rules of pronunciation for the English language.

6. R. C. CAWTHORN 1980 Internal report.
Automatic rule production.