Acquiring Pronunciation Data for a Placenames Lexicon in a Less-Resourced Language

Briony Williams, Rhys James Jones
Language Technologies Unit, Canolfan Bedwyr, Bangor University, UK
Canolfan Bedwyr, Normal Site, Bangor University, Gwynedd LL57 2PX, UK
E-mail: b.williams@bangor.ac.uk, r.j.jones@bangor.ac.uk

Abstract

A new procedure is described for generating pronunciations for a dictionary of place-names in a less-resourced language (Welsh, spoken in Wales, UK). The method is suitable for use in a situation where there is a lack of skilled phoneticians with expertise in the language, but where there are native speakers available, as well as a text-to-speech synthesiser for the language. The lack of skilled phoneticians will make it impossible to carry out direct editing of pronunciations, and so a method has been devised that makes it possible for non-phonetician native speakers to edit pronunciations without knowledge of the phonology of the language. The key advance in this method is the use of "re-spelling" to indicate pronunciation in a linguistically-naïve fashion on the part of the non-specialist native speaker. The "re-spelled" forms of placenames are used to drive a set of specially-adapted letter-to-sound rules, which generate the pronunciations desired. The speech synthesiser is used to provide audio feedback to the native speaker editor for purposes of verification. A graphical user interface acts as the link between the database, the speech synthesiser and the native speaker editor. This method has been used successfully to generate pronunciations for placenames in Wales.

1. Background

In the development of speech technology applications for a language, a pronunciation lexicon for placenames can be an important resource. However, producing it is far from straightforward. This is particularly the case for a less-resourced language (such as Welsh) where there may be no existing printed pronunciation dictionary of placenames, and where the supply of expert phoneticians is severely restricted. This means that existing methods of acquiring pronunciation data cannot be used, and so a new method has been devised here.

The work reported here is part of a project developing a pronunciation lexicon for Welsh placenames. No such lexicon currently exists, and it is needed by several classes of user: broadcasters, Welsh learners, tourists, native speakers, and users of screen-reading software in Welsh. It is being produced using a Welsh phoneset, but could be mapped to a British English (RP) phoneset if required.

Welsh is a Celtic language spoken mainly in Wales, in the UK, by about 582,000 people, of whom 458,000 speak it fluently. After a long period of decline, Welsh is beginning to increase in usage among speakers and in more contexts. A project by the present team to develop a Windows-based Welsh text-to-speech (TTS) synthesiser (Williams et al., 2006) gave rise to the first Welsh screen-reader. The work reported here is an extension of that work, which was developed within the "Festival" TTS framework (Taylor et al., 1998).

Welsh has particular features which are relevant:

a) The correspondence between Welsh orthography and pronunciation is very regular, so letter-to-sound (LTS) rules will give good results for standard Welsh words (though not always for placenames).

b) Welsh text frequently contains English words (so-called "code-switching"), and placenames in Wales could be either in Welsh or in English: many places have names in both languages. However, Welsh LTS rules give very bad results for English input words, and so are not sufficient on their own.

2 The place of the pronunciation lexicon in a TTS system

A lexicon of pronunciations, possibly including part of speech information as well, lies at the core of a TTS system. It is the first resource to be consulted by the system when converting an input string into its phonemic representation. Only if the input word does not appear in the lexicon will the system fall back to LTS rules. During development, the pronunciations in this general lexicon may be produced manually, automatically (by LTS rules), or semi-automatically (with hand-editing after LTS rules).

A recent paper comments, with regard to pronunciation lexicons for speech recognition, "Creation of pronunciation lexicons for speech recognition is widely acknowledged to be an important aspect of system development, but is it rarely addressed in detail. This is probably because the lexicons are often manually created and make use of knowledge and expertise that is difficult to codify" (Lamel and Adda, 1996). This paper attempts to go some way towards rectifying this lack of discussion.

2.1 General lexicon versus placename lexicon

In the case of a lexicon of placenames, the task of producing this lexicon differs from the task of producing a lexicon of general vocabulary. This is because the input data for a placename lexicon shows a much greater degree of pronunciation variability than does the data for a general lexicon. The reasons for this are:

a) A list of placenames will include many English items, which follow different pronunciation rules from the Welsh placenames (the two languages are from different language families within Indo-European).

b) Placenames, even in the same language, tend to show greater variability than general vocabulary (even if only in their stress patterns, in the case of Welsh).

This means that even when a set of LTS rules already exists, these rules will not necessarily be of great help when generating placename pronunciations.

2.2 Methods of lexicon production

The options for creating the placename pronunciations are the same as for the general vocabulary, as follows:

2.2.1 Manual

Each pronunciation is entered as phoneme symbols by an expert phonetician who is familiar with the language. This approach, while the most accurate, is also highly time-consuming. For a less-resourced language such as Welsh, it is not practical. This is due to the extreme shortage of phoneticians with the necessary expertise who are also familiar with speech technology. This situation will arise in other less-resourced languages as well.

2.2.2 Semi-automatic

A set of LTS rules outputs a string of phonemes, which is then hand-edited by a phonetician. This method requires a little less time and effort. However, it still suffers from the same objection as above, since it requires an expert to read and edit phoneme strings.

2.2.3 Automatic

In theory, it would be possible to run LTS rules over the input placenames without editing the output. This method is very fast and does not require human experts (once the rules are written). However, the output is certain to be much less accurate than the other methods, and so is not suitable for a dictionary that is to be the "gold standard".

2.2.4 Insufficiency of these methods

None of these methods are satisfactory for placenames in less-resourced languages, because:

a) Time and funding are in very short supply.

b) There are not sufficient trained phoneticians.

c) The input data contains items from two different languages, and with very irregular pronunciation.

Therefore an alternative method was devised for vocabulary which has very irregular pronunciation, such as placenames.

3 A new method

The existing Welsh resources available to the project comprised the following:

a) A list of placenames in Wales (in Welsh and English) from the Ordnance Survey and other sources.

b) A basic diphone-based TTS synthesis system for Welsh, including a manually-built set of LTS rules based on those in Williams (1994).

c) A database infrastructure on a shared server.

d) Native speakers without special phonetic knowledge.

LTS rules were available, but only one phonetician. The timescale was also very short due to reasons of funding. This situation will be common for any less-resourced language. Hence it was necessary to find a way to enable linguistically-naïve speakers to input pronunciations without knowledge of phonetics.

3.1 Crucial prerequisites for the method

The key innovations were as follows:

a) Welsh has the phenomenon of "re-spelling", where English words are spelled in Welsh orthography. This is seen in Welsh online forums as a jocular mode of spelling English words: e.g., "reffarî" for English "referee", or "lyfli jyb" for English "lovely job" (pronounced in a Welsh accent of English as /ləvli dʒəb/). Even linguistically-naïve native speakers are able to use "re-spelling" to indicate pronunciation.

b) An existing Welsh TTS system (developed by the present team) was available, running on MS-Windows under MSAPI. This could be used to give audio feedback on pronunciation where the user lacked the expertise to edit strings of phonemes.

3.2 Editing interface

A graphical editing interface was developed, running under Windows XP. Its logical structure was as follows (see flowchart in Fig. 1 below):

a) Input words in plain text format are extracted from the database, running on a shared server.

b) When selected by the user, the word is passed through the TTS system, and the TTS output is played out.

c) If the pronunciation is acceptable to the user, the system saves the orthographic form of the word.

d) If the pronunciation is not acceptable, the (non-linguist) user inputs a re-spelled version of the orthography. This is then passed through the TTS system and the result played back.

e) The process is repeated if needed for alternative re-spellings of the word, over all words in the input.

f) Finally, the saved orthographic (or re-spelled) forms of the input words are passed through the LTS rules in batch mode to output the desired strings of phonemes.

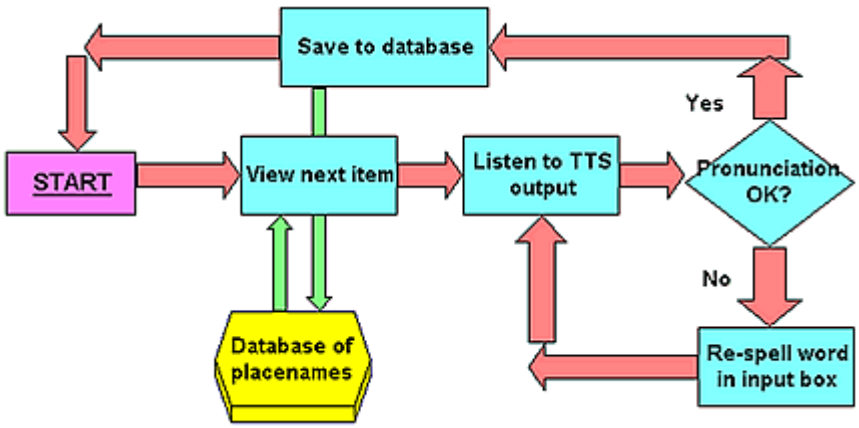

Fig 1: Flowchart showing operation of the editing interface

Modifications were made after initial testing, as follows:

a) The method needed a way of specifying vowels that were found only in English placenames (not in Welsh placenames). This was done by specifying certain symbol combinations to have certain pronunciations, using example words to guide the user.

b) A "spell-out" mode was implemented, whereby letter-names are pronounced individually (as in acronyms).

c) The vowel-lengthening function of Welsh diacritics was split from their stress-assigning function (for which an alternative and more direct representation was devised).

The user view of the editing software is shown in Fig. 2, with annotations, showing a typical English placename and its representation in the re-spelling system.

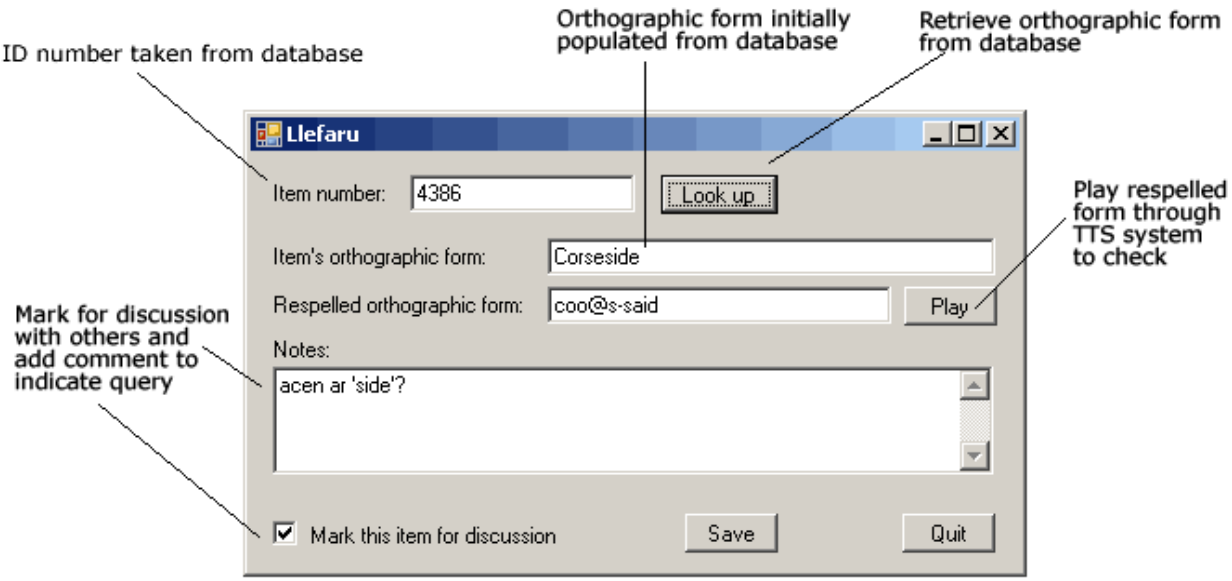

Fig. 2: Editing interface showing input fields and functions

3.3 Advantages and disadvantages

When developing lexicons with a minimum of resources, the method is an ideal compromise between available resources and development time. Manual editing is restricted to areas where it is most needed, as follows:

a) The general vocabulary shows a better correspondence between sound and spelling. Hence it is covered by LTS rules, which are used to create the main lexicon with little or no manual intervention.

b) Specialised vocabulary (e.g. placenames) shows more irregularity, and hence needs manual editing.

c) Given insufficient phoneticians, the manual editing must be done by non-specialist native speakers.

d) This can be made possible by harnessing re-spelling for foreign and irregular words, plus TTS output.

e) This protocol minimises both the manual intervention needed, and the amount of training needed.

These advantages are offset by certain disadvantages:

a) Significant time was needed to fine-tune the LTS rules for use even for respelled English words (the initial form of the rules handled only Welsh input).
b) Significant time had been needed to build the original LTS rules by hand (i.e. not statistically). This is only possible where an expert phonetician is available.
c) Minor modifications were needed to the TTS system, in order to emphasise syllable stress in the audio, and to enable the transcriber to distinguish the length of monophthongs more easily. The ease of making these modifications may depend on synthesis architecture, the voice itself and the language of the lexicon.

4 Progress and possible other uses

An input list of 5400 placenames has been processed according to this method. After initial adjustments as specified above, the method proved easy to use. It is anticipated that it will be downloadable free of charge for use by other researchers in less-resourced languages.

The method could be adapted for use in the development of a general lexicon for TTS systems in less-resourced languages, in situations where a phonetician is not available but native speakers are available. In this case, the procedure would be an iterative one, as follows:

a) A very small seed lexicon is edited by hand to provide pronunciations.
b) This lexicon is used to train a first set of statistical LTS rules.
c) These rules are used to produce a larger lexicon, and the remaining TTS components are developed.
d) Any changes necessary for the pronunciation transcription work are made to the TTS system.
e) To hand-edit the larger lexicon, the developers call on linguistically-naïve native speakers to use the method outlined above to edit the pronunciations.
f) This larger lexicon is used to train a more accurate set of LTS rules.
g) The cycle can be repeated as many times as desired.

5 References

Lamel, L. & Adda, G. (1996). On Designing Pronunciation Lexicons for Large Vocabulary, Continuous Speech Recognition. In: *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, vol. 1, pp. 6-9, 1996.

Taylor, P.A., Black A.W., and Caley, R.J. (1998) “The architecture of the Festival speech synthesis system”. In: Proceedings of the Third International Workshop on Speech Synthesis, Sydney, Australia, November 1998. Also see: http://www.cstr.ed.ac.uk/projects/festival/

Williams, B. (1994) Welsh letter-to-sound rules: Rewrite rules and two-level rules compared. *Computer Speech and Language*, vol. 8 (1994): 261-277.

Williams, B., Jones, R.J., & Uemlianin, I. (2006) "Tools and resources for speech synthesis arising from a Welsh TTS project". In: *Proceedings of the Language Resources & Evaluation Conference (LREC)*, Genoa, Italy, 2006.