Speech Technology for Minority Languages: the Case of Irish (Gaelic)

Ailbhe Ní Chasaide, John Wogan, Brian Ó Raghallaigh,
Áine Ní Bhriain, Eric Zoerner, Harald Berthelsen and Christer Gobl

Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences,
Trinity College Dublin, Ireland

{anichsid, woganj, oraghalb, anibhri, zoernee, berthelh, cegobl}@tcd.ie

10.21437/Interspeech.2006-46

Abstract

The development of speech technology could play an important role in the maintenance and preservation of minority languages, especially where the population of native speakers are dwindling. This paper outlines the efforts within the WISPR project, to develop annotated spoken corpora along with some of the prerequisites for the synthesis of Irish (Gaelic). It details the particular challenges that have confronted us as well as the strategies adopted to overcome them. It highlights the need for gearing our methodologies to these constraints and to maximise the reusability of resources. Our long-term goal is not only to develop these resources for Irish, but also, in parallel, to develop methodologies that will enable the technology to be flexible and suitable to the envisaged end users, e.g., more flexible kinds of synthesisers, with expressive capabilities and multiple voices, including children's. It is therefore a major consideration to develop resources in such a way that they are in some sense independent of any single methodology (unit selection vs. other modalities for synthesis development).

Index Terms: Irish Gaelic, text-to-speech, minority, endangered, letter-to-sound, lexicon

1. Introduction

Speech technology has a particularly crucial role to play in the maintenance and the preservation of languages whose native speaker population is diminishing. Unfortunately, given the lack of commercial incentive, these languages are lagging further and further behind in technology development. Furthermore, as this technology is becoming increasingly crucial for education and access, particularly for people with disabilities, speakers of minority languages are becoming particularly disadvantaged.

Although the know-how for developing these technologies exists, these methodologies on their own cannot deliver without specific prerequisites. For example, to develop a text-to-speech system for a language such as Irish, it is not simply enough to follow the step-by-step guidelines of, say, the Festival manual [1], one must also have in place, or develop, basic analyses of the language and related resources such as pronunciation dictionaries, suitable spoken corpora with appropriate annotations, a considerable prior linguistic knowledge of the segmental and prosodic system of the language. Providing these resources may not be a straightforward matter.

In this paper we describe our attempts to develop the speech resources required to enable the synthesis of Irish. We focus particularly on some of the difficulties confronted, and on the solutions adopted, in the hope that this experience may be helpful to other minority language groups with similar aspirations. This research has been carried out as a Welsh-Irish initiative, the project WISPR [2], funded by the EU Interreg IIIA programme. Parallel research has been carried out on Welsh, but the materials presented here pertain to the research on Irish.

In the short term, we are aiming at producing facilities according to the currently dominant technology, i.e. non-uniform unit selection concatenative synthesis, based on a large corpus of read speech recorded from a single speaker. The development to date has been carried out using the Multisyn voice in Festival [1] and the Edinburgh Speech Tools [3].

In the longer term, we aim for a flexible technology which is adapted to the user's needs: thus for example, we need not only a synthesis system which will 'speak' Irish, but also eventually, one which delivers the 'right' voice, e.g., a child voice in the right dialect/accent, with expressive capabilities. Some of our current research is in fact directed at voice variation and how expressive voices might be achieved in synthesis [4, 5, 6, 7]. Thus, it is a consideration that the resources we develop are not too tied to a single technology/methodology, but enable the harnessing of our and others' future research, geared towards such flexible voices and reusable resources.

2. Specific challenges for Irish

There are specific issues to be considered and difficulties to be overcome in order to provide for annotated corpora and speech synthesis of Irish, many of which would pertain to other minority languages.

Choosing a dialect. One of the first issues to arise was to decide which dialect of Irish to work with. There is no standard dialect of Modern Irish. Despite a long literary tradition, and a written standard from as early as the 7th century, many centuries of colonisation and the collapse of the Gaelic social order in the 17th century resulted in the decline of a language which was increasingly associated with impoverished peasants. The situation today is that the remaining communities who use Irish as a first language are scattered in the more remote regions of the Western seaboard. There are three main dialects corresponding broadly to the provinces of Ulster (the Donegal dialect), Connaught (Mayo, Connemara and the Aran Islands) and Munster (Kerry and Cork) illustrated in Fig. 1. Effectively, what is needed is a multi-dialect corpus, multi-dialect synthesis, etc. The dialect chosen in the WISPR project is that of Gaoth Dobhair in Donegal. While working on this dialect we were from the outset conscious of the need to adopt strategies that will maximally facilitate similar developments for further dialects.

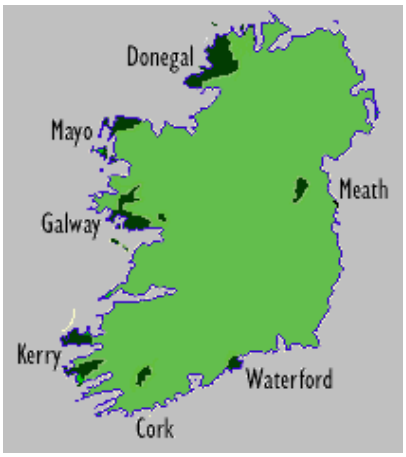

Figure 1. *Map showing in black, the main regions where Irish is spoken.*

Linguistic issues. The complexity of the Irish sound system poses an additional challenge. The consonantal system is particularly rich, involving a contrast of palatalised and velarised segments. The inventory of consonantal phonemes for Gaoth Dobhair Irish is illustrated in Table 1 [8]. Note that this table does not include the voiceless liquids and nasals: although linguists debate their phonemic status, they would undoubtedly need to be incorporated in the corpus for the purposes of synthesis. Likewise one would need to incorporate a small group of alveolar and postalveolar sounds of English, which have entered the system through borrowed words. We calculate that in total, one would need to cater for up to 68 segments.

Although from a linguist's perspective the main complexity is in the consonantal system, in fact the secondary articulation of the consonants has major effects on the realisation of an adjacent vowel: when a palatalised consonant occurs in the vicinity of a back vowel, or when a velarised consonant occurs with a front vowel, long diphthongal glides tend to arise. Thus, for example, the phoneme /i:/ may be realised as [i], [ɤi], [iɤ] or [iɤi] depending on the context. This has implications for the coverage of contextual variants one needs to ensure in the corpora, but may provide a particular challenge for the concatenation process in synthesis.

The orthographic system of Irish is also complex, reflecting its archaic origins. The opacity of the grapheme-to-phoneme correspondences can be illustrated by the rather extreme example of the phrase *Ní bhfaighfidh* 'will not get' which corresponds to the sound string [n̪ʲi: wi:].

Code switching is also common between Irish and English, given that virtually all Irish speakers also speak English. The likelihood of English words and phrases occurring in many kinds of Irish texts is high. For this reason, when recording a corpus to produce a synthetic voice, it was important to provide some possibility of an English voice, using the same speaker.

Resources. To construct a text-to-speech system one needs to have or require a fully transcribed spoken corpus. As this was not already available for Irish, it was the major focus of the WISPR project. The ideal corpus would involve optimised text that ensured good coverage of all important sound variants in the their different prosodic contexts. It would be helpful to know which sounds and sound combinations are frequent or rare, but again, this information is not yet available for Irish. So the process is circular: the tools we need to construct an annotated corpus will only become available when we have developed the corpus, and some of the annotation tools.

Table 1. *Inventory of consonantal phonemes for the Gaoth Dobhair dialect of Irish.*

	Labial	Dental	Alveolar	Alveolo-palatal	Palatal	Velar	Glottal
Plosive	pˠ bˠ pʲ bʲ	t̪ˠ d̪ˠ		t̪ʲ d̪ʲ	c ɟ	k g	
Fricative/ Approximant	fˠ w fʲ vʲ		sˠ	ɕ	ç j	x ɣ	h
Nasal	mˠ mʲ	n̪ˠ	n	n̪ʲ	ɲ	ŋ	
Tap			ɾˠ ɾʲ				
Lateral Approximant		l̪ˠ	l	l̪ʲ			

A good pronunciation lexicon is crucial to allow the development of automatic segmentation. Although a pronunciation dictionary of Irish does exist, *An Foclóir Póca* [9], it does not reflect the speech of any one spoken dialect, but represents rather an attempt at providing official standardised forms, that compromise among the dialects [10]. It was thus necessary to build a pronunciation lexicon adapted to the dialect to be used for the recorded corpus and the eventual synthesiser.

Prior phonetic/phonological analyses are important to allow transcription of the corpus. In the case of Irish, there are many phonetic descriptions of individual dialects. These are an excellent resource, but they are not always ideal for the task at hand, some of them being quite dated, and directed at the speech of the oldest members of the community. A major difficulty is that they deal almost entirely with the segmental level: there has been almost no coverage of the intonation and timing structures of Irish. This is a gap that is being targeted in our concurrent project *Prosody of Irish Dialects* [11]. This project tackles all the main dialects, and it is likely that the output of that project will go some way to facilitating our medium-term goal of good multi-dialect synthesis.

3. Developing annotated spoken corpora

Two corpora were collected, a unit-selection corpus and an enriched diphone corpus. The primary (unit selection) corpus was based on 15 hours of recorded speech using a female speaker of the Gaoth Dobhair dialect of Donegal. As just mentioned, while in principle one would have wished to work with a well-designed compact corpus that provides maximum coverage of sound sequences in a minimum of recording time, this was not possible for Irish. Although we can calculate how many sounds or diphones we need to cover, in the absence of prior annotated data, we cannot calculate their frequency and

identify which are rare. For the same reasons, pruning the corpus at the outset was not a possibility either: to do so one would need to have in place letter-to-sound rules and automatic alignment facilities. Thus, proceeding with a large recording seemed advisable, given the complexity of the sound system and the need to ensure reasonable coverage of sound variants, and sound combinations in various contexts.

It was important that the texts for the large read corpus be based on writings specific to that dialect. If one were to include texts based on other dialects, one would run the risk of the informant switching between dialect forms and introduce inconsistencies into the corpus. Suitable texts for this dialect were by and large not available in electronic form, and therefore materials were scanned. The novels of the Donegal author Séamus Mac Grianna (Máire) were the primary source. As there is no optical character recogniser for Irish, the scanning resulted in numerous errors, which were hand-corrected.

The second corpus recorded was an extended diphone corpus. The decision record a diphone corpus was based on a number of considerations. First of all, it will enable straightforward development of a diphone synthesiser at a future date. Although such synthesisers do not achieve the naturalness of good unit selection synthesis, they do often provide a more consistent quality and can be preferable for certain applications. This type of synthesis may also be more suited to certain research avenues we will want to pursue in the future, aimed at voice adaptation. The development of a diphone synthesiser will of course require more extensive basic research in the future, to model the intonation and temporal patterns, and as mentioned, some of this work is already underway. The large unit selection corpus will of course be a useful resource towards this end.

A second reason for recording a diphone corpus was that, by incorporating it into the unit selection corpus, we ensure that there is complete coverage of all occurring sound sequences. As the diphone corpus has been recorded using the same speaker and recording conditions as the unit selection corpus, it can simply serve as an extension, and safety net for it.

There is a symbiotic relationship between the two corpora. On the one hand, the diphone corpus complements the unit selection one, hopefully ensuring coverage, while the unit selection corpus will allow us to extend the diphone corpus further by extracting for example, more variants, in more prosodic contexts.

While a diphone recording typically includes all possible combinations of phones in a language, the set recorded in WISPR was substantially enriched to take account of the complexity of the sound system of Irish. It includes across word boundary diphones, Consonant1-Vowel-Consonant2 sequences, where every combination of palatalised/velarised C1 and C2 were elicited, syllables that included clusters of the form CCV, CCCV, VCC. Although a minimalist approach would suggest 55 phonemes and about 3,000 diphones for Irish, the enriched diphone corpus amounted to over 11,500 units.

4. Developing a lexicon

The only available pronunciation lexicon was the pocket dictionary *An Foclóir Póca* [9] containing 15,000 entries. As mentioned, the entries do not reflect any single dialect, but rather an attempt at providing official standard *Lárchanúint* forms that are a compromise between the existing dialects [10]. A first task therefore was to find a way to adapt the pronunciations of *An Foclóir Póca* to the specific dialect of Donegal, chosen for the corpus and synthesis development. This work has drawn on ideas in [12].

To begin with a short corpus (20 minutes) was phonetically transcribed by hand. The words (orthographic form + phonetic forms) were extracted to produce a mini-Donegal lexicon. The lexical items from this corpus were compared to the forms in *An Foclóir Póca*, to develop *sound-to-sound* rules, using the WAGON tool [3]. Note that this tool is normally used to generate statistically based letter-to-sound rules, but in this case was used to map between two sets of phonetic forms. The output rules were then applied to *An Foclóir Póca*, to produce a draft dialect-specific version. The process was carried out in stages, beginning with the most common 500 headwords. Predicted pronunciations were then hand corrected to ensure that they conformed to attested Donegal forms. Once corrected, they were added to the Donegal lexicon. This process eventually yielded a Donegal-adapted version of *An Foclóir Póca*, one that included 1,000 additional words from the 20-minute corpus.

At this point, lexicon development proceeded in tandem with the automatic segmentation of the unit selection corpus and the development of letter-to-sound rules. In order to automatically segment a corpus by forced alignment, the lexicon must contain all the words found in that corpus. As portions of the unit selection corpus were to be segmented, words not hitherto in the lexicon had to be added along with their pronunciations. This was done by applying the Donegal letter-to-sound rules to the new lexical entries. Before new entries were added they were hand checked for accuracy. This process was reiterated until all the words of the corpus were entered. This yielded a final count of 24,000 words.

5. Developing letter-to-sound rules

Two quite distinct approaches were adopted. Firstly, statistically based rules were generated, again using the WAGON tool [3], and the Donegal-adapted lexicon. The process was also an iterative one: with successive versions of the lexicon, new letter-to-sound rules could be generated. This approach was only moderately successful: there was a rather high error rate in predicting phonetic forms. It may be that despite working well for languages such as English, it is not entirely suited to the orthography of Irish. A further difficulty with this approach is that the rules are not accessible. This makes it difficult to eliminate inconsistencies.

Given our intention to develop in the future multi-dialect corpora and synthetic voices, it would be highly desirable to have accessible rules and ways of differentiating between those that are universal to Irish, and those that are specific to an individual dialect. For this reason, the second approach adopted was the development of handwritten letter-to-sound rules encoded for use within Festival. These rules were based in the first instance on *An Foclóir Póca*, with subsequent removal, reordering, and addition of rules. The results of this approach are considerably better than what was achieved automatically. Consequently, the demo synthesis voice that has been assembled uses only the handwritten rules.

Ultimately, what we aspire to with our letter-to-sound rules is a two level description. The first would be a mapping from the orthography to a 'common core' phonological representa-

tion, characterising those phonological structures that are common to the dialects. This draws on ideas proposed by Ó Murchú [13] and the philosophy that underlies the *Lárchanúint* proposals of Ó Baoill. The second level of the description would map from these common forms to the dialect specific realisations. This would not only be a neat way of capitalising on resources when adding new dialects, but promises to offer new insights into the sound structure of modern Irish dialects. Furthermore, given that the orthographic forms are essentially archaic representations of the pronunciation of Irish, this work should have a diachronic interest as well, as it should yield one model of the evolution of the modern dialects.

6. Inclusion of an Irish-English voice

Code switching is a normal feature of modern Irish. Virtually all speakers are bilingual, and live in a world where English overwhelmingly dominates many aspects of their lives. Although less prevalent in texts, it is nonetheless very frequent, especially in texts which purport to be representations of true daily conversational styles.

It was therefore considered important to enable our first Irish synthesiser to be able to 'speak' the words and phrases of English that will appear in some of the texts. The strategy adopted was to build a parallel Irish-English synthesiser, using the Arctic corpus [14], which is a compact corpus designed to yield coverage of the phonemes of English. This has been extremely useful in allowing us to rapidly put together a parallel Irish-English synthesiser.

7. Conclusions

Currently there is an annotated 15-hour corpus, an annotated extended diphone corpus, and an initial, demonstration Irish synthetic voice. These resources are far from complete: many aspects of a full text-to-speech system are yet to be done, e.g., on tokenisation, prosody modelling, etc. And although the demonstration voice is far from perfect, it demonstrated the power of the recorded corpus, while highlighting problems with the annotations, which will require considerable work. Nonetheless this is an important first step, the beginning we hope of an extended programme to provide annotated corpora, text-to-speech and other speech technology facilities for the dialects of Irish. By extending the Welsh collaborations fostered by WISPR, we aspire to jointly extend to collaborations with other Celtic languages, including Scottish Gaelic (a near relative of Irish) as well as Breton and Cornish (near relatives of Welsh). It is also our hope that strategies we evolve to deal with the linguistic and resource difficulties, and the strategies to facilitate multi-dialect and multi-language synthesis will be of use to others who share our concerns for the future of their language.

8. Acknowledgments

This research was funded by the EU Interreg IIIA Community Initiative Programme. The work has also benefited from interactions with the EU-funded Humaine Network of Excellence on Emotion, and from the project *Prosody of Irish Dialects*, which is funded by the Irish Research Council for the Humanities and Social Sciences.

9. References

[1] http://www.cstr.ed.ac.uk/projects/festival

[2] Welsh and Irish Speech Processing Resources, EU-funded Interrreg IIIA project, 2004-2005, http//:www.tcd.ie/CLCS /phonetics/projects/prosody.html

[3] http://www.cstr.ed.ac.uk/projects/speech_tools

[4] Yanushevskaya, I., Gobl, C., and Ní Chasaide, A., "Voice quality and f_0 cues for affect expression: implications for synthesis", *Proceedings of the 9th European Conference on Speech Communication and Technology, INTERSPEECH 2005*, Lisbon, 1849-1852, 2005.

[5] Gobl, C. and Ní Chasaide, A., "The role of voice quality in communicating emotion, mood and attitude", *Speech Communication*, Vol. 40, 189-212, 2003.

[6] Ní Chasaide, A. and Gobl, C., "Voice Quality and the Synthesis of Affect," in E. Keller, G. Bailly, A. Monaghan, J. Terken and M. Huckvale (eds.) *Improvements in Speech Synthesis,* Wiley and Sons, 252-263.

[7] Ní Chasaide, A. and Gobl, C., "Decomposing linguistic and affective components of phonatory quality," *Proceedings of the 8th International Conference on Spoken Language Processing, INTERSPEECH 2004*, Jeju Island, Korea, Vol. 2, 901-904, 2004.

[8] Ní Chasaide, A., "Irish", in *The Handbook of the International Phonetic Association*, Cambridge University Press, Cambridge, 111–116, 1999.

[9] Rialtas na hÉireann, *An Foclóir Póca*, An Gúm, Dublin, 1986.

[10] Ó Baoill, D. P., Lárchanúint donGhaeilge. *Institiúid Teangeolaíochta Éireann*, Dublin, 1986.

[11] Prosody of Irish Dialects: the use of intonation, rhythm, voice quality for linguistic and paralinguistic signalling, IRCHSS-funded project, 2003-2006, http://www.tcd.ie/ CLCS/phonetics/projects/prosody.html

[12] Maskey, S. R., Black, A. W., and Tomokyo, L. M., "Bootstrapping phonetic lexicons for new languages", *Proceedings of the 8th International Conference on Spoken Language Processing, INTERSPEECH 2004*, Jeju Island, Korea, 69-72, 2004.

[13] Ó Murchú, M., "Common core and underlying forms, *ÉRIÚ*, 21: 42-75, 1969.

[14] http://festvox.org/cmu_arctic/cmu