

Reverse-engineering the morphophonology of Gaelic vowels from orthography: Language planning to linguistic theory, and back again

Mark McConville¹

This paper investigates the morphophonology and orthography of Scottish Gaelic vowels, with a view to investigating exactly how many contrastive short stressed vowels there are in the language, and also to considering the various implications of this problem for ongoing language planning efforts.

I start (§1) by introducing the concept of ‘deep’ phonology, using the better understood topic of Gaelic consonant morphophonemics by way of example. I demonstrate (§2) that there turns out to be a very close connection between the morphophonological analysis of the consonants and the traditional Gaelic orthography, the latter being best characterised as a ‘deep’ phonographic system. I then turn (§3) to whether we can apply this insight to the study of Gaelic vowels – can we learn anything about the contrastive short stressed vowel inventory from the way they are represented in writing, in effect reverse-engineering the morphophonemic structure of Gaelic vowels from their traditional orthographic representation?

My methodology (§4) involved the creation of a gold-standard list of 2,264 Gaelic monosyllabic nouns, verbs and adjectives from a combination of electronic and paper-based dictionaries. A thorough empirical analysis of this data (§5-8) suggests that there are just *three* basic short stressed vowel morphophonemes – high, mid and low – along with a few lexically marked rarities after slender onsets. In particular, phonological roundedness and front-ness are almost always predictable from the broad/slender value of the onset and coda.

¹ This work was supported by the Soillse national research network for Gaelic language and culture. I am particularly grateful to Professor Roibeard Ó Maolalaigh who provided detailed comments on an earlier version.

I conclude (§9) with a discussion of the implications of this analysis of short stressed vowels for Gaelic language planning, especially the potential of orthographic reform to support language acquisition.

1 Some background:

Morphophonology of Gaelic consonants

Gorrie (2011) argues that, in order to describe the inflectional patterns of attributive adjectives in Scottish Gaelic within the framework of Distributed Morphology² (Halle and Marantz 1993), it is necessary to posit a level of ‘morphophonemic’ representation, distinct from both phonemic representation on the one hand and morphosyntactic representation on the other. He demonstrates that the productive morphosyntactic processes of initial lenition and final slenderisation of Gaelic adjectives cannot be interpreted straightforwardly either in terms of the grammatical categories of case, gender and number that they encode, or in terms of the phonological features such as [continuant] and [palatal] that they are realised with, hence the need for two binary morphophonemic features, [lenited] and [slender], to mediate between the two.

Implicit within Gorrie’s (2011) analysis is a deep consonant inventory for Gaelic that can be summarised as follows:

- There are twelve basic consonant types in Gaelic, one for each of the consonant graphemes in traditional Gaelic orthography: B, C, D, F, G, L, M, N, P, R, S and T.³
- Every consonant is either fortis (e.g.) or lenis (e.g. <bh>).
- Every consonant is either broad or slender.

This gives a complete inventory of 48 fully specified consonant morphophonemes in Gaelic. Morphophonemic representations are stored in the lexicon and manipulated by the morphosyntactic component, and at the grammar/phonology interface are interpreted (or ‘spelled out’) as sequences of phonemes.

Interestingly, at the purely phonemic, surface level, Gaelic appears to make many fewer consonantal distinctions than at the morphophonemic, deep level – both Ternes (2006) and Bosch (2010) posit an inventory of

² Distributed Morphology holds that the processes of word-formation are not particular to the lexicon, but are distributed among other components of the grammar, especially the morphosyntactic component. However, the argument presented in this paper does not depend on these assumptions about the architecture of the grammar.

³ I do not include <h> as a basic consonant grapheme, regarding it rather as an ‘auxiliary’ grapheme (see below).

around 30 consonant phonemes in modern Gaelic. This discrepancy is accounted for by: (1) cases of syncretism in the realisation of consonant morphophonemes, e.g. both broad lenis S and broad lenis T are typically realised as /h/, and broad lenis P and fortis F are both realised as /f/; (2) cases where a single morphophonemic unit is realised as a sequence of phonemes, e.g. slender fortis B, F, M and P (i.e. the labials) are generally realised as /bj/, /fj/, /mj/ and /pj/ in initial position⁴; and (3) cases of null realisation, e.g. broad lenis F is never realised phonemically.

Finally, it should be noted that the relation between the morphophonemic and phonemic levels of representation cannot be described simply in terms of a one-to-one mapping involving individual morphophonemes. Rather, the realisation of a consonant morphophoneme is often highly context-sensitive. To give just one example, when broad lenis B occurs as the coda to a syllable with a slender onset and a low vowel nucleus, then the entire rhyme of the syllable is realised simply as the vowel phoneme /ɔ/, e.g. *treabh* (to plough) is generally realised phonemically as /trɔ/, and *feabhas* (goodness) as /fjɔ.əs/ (Bauer 2011: 515).

2 Gaelic orthography as a deep phonographic system

Traditional Gaelic orthography is generally assumed to be phonographic in nature, where graphemes are associated directly with sounds, independent of meaning. This accounts for the following general features of Gaelic orthography: (a) if you know the written form of a Gaelic word, then you can predict its spoken form with a very high degree of confidence, even if you do not know what the word means; (b) similarly, if you know the spoken form of a Gaelic word, then you can predict its written form with a very high degree of confidence, again without needing to know what it means; and (c) there are almost no heterographic homophones (words which are spelled differently but pronounced the same) or heterophonic homographs (words which are pronounced differently but spelled the same) in Gaelic orthography.

When one contrasts: (a) the relation between consonant phonemes and graphemes in Gaelic, with (b) the relation between consonant morphophonemes and graphemes, it is clear that it is the latter that is by far the more transparent. In other words, traditional Gaelic orthography is a

⁴ This analysis used to be controversial, cf. Oftedal (1965), MacAulay (1968), possibly because of a confusion between the phonemic and morphophonemic levels of sound structure.

‘deep’ phonographic system, rather than a ‘surface’ one, since graphemes are directly associated with Gorrie’s morphophonemic features, and only indirectly with phonemes or phonological features, at least as far as the consonants are concerned. More precisely:

- There is a one-to-one correspondence between the twelve basic consonant morphophoneme types B, C, D, F, G, L, M, N, P, R, S and T and the twelve core graphemes , <c>, <d>, <f>, <g>, <l>, <m>, <n>, <p>, <r>, <s> and <t>.
- The morphophonemic feature [lenis/fortis] is represented by the presence or absence of the auxiliary grapheme <h> immediately following the core grapheme representing the basic consonant type.⁵
- The morphophonemic feature [broad/slender] is represented by the choice of auxiliary vowel grapheme immediately preceding and/or following the consonant graphemes. The graphemes <a>, <o> and <u> denote a broad consonant, whilst <e> and <i> denote a slender consonant.

In effect, the orthography of Gaelic consonants is a systematic mapping from the 48 consonant morphophonemes to a range of several hundred di-, tri- and tetra-graphs. For example, here are the theoretically possible orthographic realisations for consonants of basic type B:⁶

- [broad fortis B]
 - initial: <ba>, <bo>, <bu>
 - final: <ab>, <ob>, <ub>
 - medial: <aba>, <abo>, <abu>, <oba>, <obo>, <obu>, <uba>, <ubo>, <ubu>
- [broad lenis B]
 - initial: <bha>, <bho>, <bhu>
 - final: <abh>, <obh>, <ubh>

⁵ Except for the sonorants L, N and R, where the lenis/fortis distinction is neutralised orthographically (though not phonemically) in initial position, and marked by the absence or presence of consonant doubling in final position, e.g. lenis <-l>, <-n>, <-r> versus fortis <-ll>, <-nn>, <-rr>. The digraphs <lh>, <n>h and <rh> are not part of standard Gaelic orthography, and neither are initial <ll->, <nn-> and <rr->, notwithstanding some earlier attempts to introduce them (Black 2010).

⁶ Not all of these occur in practice, due to independent restrictions on the vowel grapheme sequences that can appear in stressed and unstressed syllables.

The morphophonology of Gaelic vowels

- medial: <abha>, <abho>, <abhu>, <obha>, <obho>, <obu>, <ubha>, <ubho>, <ubu>
- [slender fortis B]
 - initial: <be>, <bi>
 - final: <eb>, <ib>
 - medial: <ebe>, <ebi>, <ibe>, <ibi>
- [slender lenis B]
 - initial: <bhe>, <bhi>
 - final: <ebh>, <ibh>
 - medial: <ebhe>, <ebhi>, <ibhe>, <ibhi>

The system gets more complicated when we take into account the full range of consonant clusters found in Gaelic words, but the basic principles remain the same.

One side-effect of this deep, morphophonemic feature-based system is that Gaelic orthography manifests a property that we might call ‘logographic transparency’, whereby all inflectional variants of a particular lexeme share a common orthographic core. Take for example the masculine common noun *tùs* (source) which is found in four different monosyllabic variants:

- *tùs* - /tu:s/ (nominative/dative singular)
- *thùs* - /hu:s/ (nominative/dative singular after leniting particles)
- *tùis* - /tu:f/ (genitive singular)
- *thùis* - /hu:f/ (genitive singular after leniting particles)

The traditional Gaelic orthography allows for the common orthographic core <t ù s> for these four forms, whereas a strictly phoneme-based, surface phonographic orthography would only allow for the vowel grapheme <ù> to be shared between them, as the phonemic representations make clear. By constructing a phonographic orthography based on deep morphophonemic features rather than surface phonemes, and by applying the use of auxiliary graphemes systematically throughout the entire consonantal system, the Gaelic scholars of the late medieval period successfully solved a genuine language planning conundrum – how to simultaneously satisfy both phonographic and logographic transparency in a language where most of the morphosyntactic ‘heavy lifting’ is carried out by consonant mutation – using just thirteen consonant and five vowel graphemes from the Latin alphabet to represent an inventory of 48 distinct consonant morphophonemes.

Aside from the realisation of individual consonants, there are at least three other phenomena which provide evidence for the claim that traditional Gaelic orthography is a deep phonographic system rather than a surface phonemic one. The first of these involves compensatory lengthening, where a short stressed vowel morphophoneme which is immediately followed by a word-final fortis sonorant is obligatorily diphthongised or lengthened during the transition from deep to surface phonology. For example, the vowels in the words *fonn* (tune), *cam* (bent), *cill* (churchyard) and *barr* (top) are realised respectively as /ɔu/, /au/, /i:/ and /a:/, rather than /ɔ/, /a/, /i/ and /a/ (Bauer 2011: 340). In the traditional orthography, this surface lengthening is generally not marked. A related phenomenon involves stressed vowels immediately followed by an *-rd*, *-rn* or *-rl* consonant cluster, which are also obligatorily lengthened, for example *bòrd* (table) and *càrn* (cairn). Again, in traditional orthography, this surface lengthening was not always marked explicitly.⁷

A second phenomenon that supports the notion that traditional Gaelic orthography is a deep phonographic system involves the ‘intrusive’ sounds which turn up in the pronunciation of certain consonant clusters, but not in the orthography: (a) the cluster *-rt* is almost always pronounced with an intrusive sybillant (typically [ʃ]), e.g. *ceart* (right), *beartach* (rich); (b) in many dialects, the same goes for *-rd* clusters, e.g. *bòrd* (table), *àrd* (high); and (c) the cluster *sr-* is frequently pronounced with an intrusive [t], e.g. *sròn* (nose), *srath* (valley) (Bauer 2011: 350). A third phenomenon illustrating the deep nature of traditional Gaelic orthography is *svarabhakti*, where consonant clusters consisting of a sonorant followed by a non-homorganic voiced obstruent or unvoiced fricative (e.g. *-lb*, *-lg*, *-rg*, *-rm*, *-nbh*, *-lbh*, *-rch-*, *-mch-*, *-ms-*) in a stressed syllable undergo a kind of epenthesis. For example, *gorm* (blue), *dearg* (red) and *Alba* (Scotland) are always pronounced with an epenthetic vowel, for example as [gɔrɔm], [dʲɛrɛk] and [aLabə] respectively (Bauer 2011: 334), and the epenthetic vowel is usually not represented in the orthography.

3 The problem of the Gaelic short vowel inventory

In the previous section, starting out from Gorrie’s (2011) independently motivated, feature-based analysis of the deep phonology (or morphophonemics) of Gaelic consonants, it was demonstrated that the

⁷ However, recent official Gaelic orthographic guidelines (SQA, 2009: part 12) have confused matters somewhat, with an implicit preference for using an accent to represent a surface-lengthened vowel before *-rr*, *-rd* and *-rn*.

The morphophonology of Gaelic vowels

traditional consonantal orthography appears to be a transparent, sophisticated and elegant reflection of the underlying morphophonemic structure of the language, rather than having any kind of direct relation with surface phonemic representation.

Turning now to the subject of Gaelic vowel sounds, most descriptions of the phonology of Gaelic (e.g. Ternes 2006) specify an inventory of *nine* short vowel phonemes which can occur as the nucleus of a stressed syllable. These are classified by Bosch (2010) along two basic axes – three values for vowel height, and a two-way distinction between front vowels and back vowels – and in addition, the back vowels are subcategorised according to two roundedness values, as seen in Table 1.

| | front | back | |
|-----------|-------|-----------|---------|
| | | unrounded | rounded |
| high | /i/ | /ɯ/ | /u/ |
| mid | /e/ | /ɤ/ | /o/ |
| lower-mid | /ɛ/ | /a/ | /ɔ/ |

Table 1 – Bosch’s (2010) inventory of short stressed vowels in modern Gaelic

However, although there is a consensus that modern Gaelic contains nine short stressed vowel phonemes as the surface level, it remains unclear exactly how many of these are actually lexically contrastive. Unlike the situation for Gaelic consonants, there has been little or no investigation into the deep phonology of Gaelic vowels, presumably because of lack of relevant data, vowel mutation (e.g. *Umlaut* and *Ablaut*) being relatively unimportant operations in the morphosyntax of the language.

The lack of clarity surrounding the Gaelic vowel inventory is implicated in three observations regarding the realisation of vowels in the Gaelic lexicon. First of all, there is evidently a lot of assimilation at work, where the quality of a vowel depends on the phonological nature of the preceding and following consonants. For example: (a) non-low back vowels tend to be rounded before broad consonants and unrounded before slender ones (Bauer 2011: 550, 580), suggesting that /u/ and /o/ are in complementary distribution with /ɯ/ and /ɤ/ respectively; and (b) non-low

front vowels never appear after broad consonantal onsets, apart from in unassimilated loanwords like *tidsear* (teacher) and *ticead* (ticket). Secondly, there is the well-known dearth of minimal pairs in Gaelic involving contrasting short vowels, and in particular the non-existence of meaningful contrasts involving front and back vowels, or rounded and unrounded vowels. Finally, there is the fact of the large amount of dialectal and ideolectal variation involving vowels, which has the effect of neutralising certain vowel distinctions in certain contexts, for example the variable realisation of the vowel in *bainne* (milk) across the Gaelic dialects – /bajə/, /bɔjə/, /bejə/, /bejə/, ... (Bauer 2010: 309). Together, these three observations suggest that the inventory of actually contrasting Gaelic short vowels may well have fewer than nine members.

Gorrie's independently motivated, feature-based analysis of the morphophonemics of Gaelic consonants turns out to be very close in many respects to what we would expect based solely on a careful analysis of the traditional Gaelic orthography. This raises a very interesting question with respect to the vowel problem – can we learn anything about the Gaelic contrastive short vowel inventory from the way they are represented in writing, and thereby compensate for the lack of relevant morphosyntactic data involving vowel mutations? More ambitiously, can we reverse-engineer the morphophonemic structure of Gaelic vowels from their traditional orthographic representation, using language planning as a window into linguistic theory?

4 Methodology - data collection

In order to investigate the orthography of Gaelic stressed vowels with a view to reverse-engineering their morphophonemics, I created a gold-standard list of Gaelic monosyllabic words from a combination of electronic and paper-based dictionaries.

The starting point for data collection was Sabhal Mòr Ostaig's online version of Malcolm MacFarlane's (1912) Gaelic dictionary for schools, *Am Briathrachan Beag*,⁸ which contains a total of 5,589 entries. I automatically harvested the 1,584 monosyllabic entries,⁹ and then went through this wordlist by hand, comparing each entry with the entries in Mark (2004). In the course of this, the following steps were undertaken:

⁸ <http://www2.smo.uhi.ac.uk/gaidhlig/faclair/bb/bb-codes.txt>

⁹ Unless otherwise specified, all text processing was done using standard Unix command line tools, like `grep` and `awk`.

- Words which are not native open-class nouns, verbs, adjectives or adverbs were removed, in order to ensure that only stressed monosyllables are in the wordlist. The words which were deleted during this step include prepositions, prepositional pronouns, pronouns, and other grammatical particles, as well as obvious recent loanwords from English.
- Where a word has a slenderised morphological variant, this was added to the wordlist, to ensure maximum coverage. For example, the slenderised (dative singular) form *làimh* of the noun *làmh* (a hand) was added.
- Vowels marked as long but which clearly exhibit compensatory lengthening were normalised to their presumed underlying short equivalents, in order to maximise consistency across the wordlist. For example, *sgùrr* (a crag) was normalised to *sgurr* based on a comparison with related words with a short vowel like *sgurrach* (rugged), but *ciùrr* (to harm) was not normalised, since related words in Mark's dictionary retain the long vowel, e.g. *ciùrramach* (harmful).
- Finally, monosyllabic nouns, verbs and adjectives which appear in Mark (2004) but not in the original list were added to the wordlist as well.

The final, hand-checked, gold-standard version of the wordlist contains 2,264 monosyllabic, open-class Gaelic words, including both citation forms and morphologically slenderised variants.

5 Basic structure of the monosyllable wordlist

The wordlist contains 37 orthographically distinct vowel sequences, which can be partitioned along two independent axes: (a) the broad/slender value of the surrounding consonants; and (b) whether the vowel is considered to be short or long (including diphthongs). The number of instances of each orthographic vowel sequence in the wordlist is presented in Table 2.¹⁰

A close analysis of this frequency distribution brings out two phonological issues which are worthy of note. First of all, since 64% of the total words in the wordlist have a broad onset, we are justified in concluding that this is the default option in the Gaelic lexicon. Words which have slender onsets are thus assumed to be lexically marked. Secondly, there are

¹⁰ The original wordlist did not contain any words with an acute accent, having been normalised to a modern, grave-only format during digitisation.

The morphophonology of Gaelic vowels

no more than six potential vowel contrasts in any context, which immediately gives us an upper bound on the size of the Gaelic stressed vowel inventory – there are no more than six contrasting short lexical vowels and six contrasting long lexical vowels.

| | S_S | | S_B | | B_B | | B_S | |
|--------------|-----|----|-----|-----|-----|-----|-----|-----|
| short | e | 2 | ea | 172 | a | 186 | ai | 158 |
| | eai | 4 | eo | 4 | o | 155 | oi | 76 |
| | ei | 92 | io | 52 | u | 80 | ui | 159 |
| | eo | 2 | iu | 9 | | | | |
| | i | 83 | | | | | | |
| | iui | 1 | | | | | | |
| long | è | 5 | èa | 9 | à | 87 | ài | 92 |
| | èi | 85 | eò | 20 | ao | 52 | aoi | 76 |
| | eò | 26 | eu | 63 | ò | 67 | òì | 53 |
| | ì | 45 | ia | 64 | ù | 61 | ùì | 50 |
| | ìù | 9 | ìo | 42 | ua | 45 | uai | 63 |
| | | | ìù | 14 | | | | |

Table 2 – Frequency distribution of orthographic vowel sequences in the wordlist

The remainder of this paper will mainly be concerned with the *short* stressed vowels, i.e. in the top half of Table 2. There is a total of 1,235 short stressed monosyllables in the wordlist, including normalised forms like *bord* (table) and *barr* (top).

6 Short vowels after broad onsets

The morphophonology of Gaelic vowels

As seen in Table 2, after a broad onset (i.e. in B_B and B_S contexts) the orthography suggests a three-way, height-based short vowel contrast, between: (a) the *high* vowels <u> and <ui>, typically pronounced as /u/ and /ui/; (b) the *mid* vowels <o> and <oi>, typically pronounced as /ɔ/ and /ɔi/; and (c) the *low* vowels <a> and <ai>, typically pronounced as /a/ and /ɛ/. The frequency distribution also suggests the following generalisations:

- There is a tendency to prefer lower vowels before a broad coda (i.e. in B_B contexts) – only 19% of broad codas are preceded by <u>, compared to 37% by <o> and 44% by <a>.
- However, there is no corresponding tendency to prefer higher vowels before a slender coda (i.e. in B_S contexts) – the proportions of slender codas preceded by <ai> and <ui> are both 40%.
- Rather, there appears to be a general tendency to avoid mid vowels, which is overridden before a broad coda by the preference for lower vowels.

One thing that the orthography does not imply is a meaning-bearing distinction between rounded and unrounded vowel phonemes after a broad onset. Rather, as suggested above, this appears to be a simple matter of surface phonological assimilation, where vowels have a tendency to be unrounded before a slender coda.

With regard to the question of minimal pairs, it was found that, of the 649 B_B and B_S contexts that appear in the wordlist:

- There are 13 three-way contrasts, e.g. *lach* (duck) - *loch* (lake) - *luch* (mouse), *mail* (prince; genitive) - *moil* (shingle; genitive) - *muil* (axis - genitive).
- There are 139 two-way contrasts, e.g. *bann* (belt) - *bonn* (medal), *gairm* (to call) - *guirm* (blue; slenderised).

In other words, although there *is* a three-way, height-based short vowel contrast after broad onsets, this does not actually carry much in the way of information, since the vast majority of words with broad onsets do not contrast with any other word in terms of their vowel nucleus.

Finally, it is also instructive to look at the morphological slenderisation patterns in words with a broad onset:

- All words in <u> slenderise to <ui>, e.g. *muc* - *muic* (pig).

The morphophonology of Gaelic vowels

- 27% of words in <o> slenderise to <oi> and 70% to <ui>, e.g. *bonn - buinn* (medal), *ros - rois* (seed). With the remainder usage varies, e.g. *dos - dois/duis* (thicket).
- 82% of words in <a> slenderise to <ai>, 6% to <oi>, and 8% to <ui>, e.g. *bàrd - bàird* (poet), *cas - cois* (foot), *clag - cluig* (bell).

Thus, when a B_B word is converted into a B_S word by means of morphological slenderisation, it appears to lose the tendency to prefer lower vowels and to gain that of avoiding mid vowels.

7 Short vowels after slender onsets

As seen in Table 2, after a slender onset the orthography suggests a much more confusing picture than is the case after a broad onset.

Before a broad coda (i.e. in S_B contexts), there are four orthographically distinct vowel sequences: <io>, <ea>, <iu> and <eo>, pronounced in most contexts as /i/, /ɛ/, /u/ and /ɔ/ respectively. However, the latter two occur much less frequently than do the first two, suggesting that they are statistical outliers in some way. To be precise:

- <iu> is found in only nine words: *fliuch* (wet), *liurc* (to wrinkle), *liut* (knack), *pliut* (a flipper), *spliug* (snot), *spliut* (webbed foot), *sriut* (tirade), *tiugh* (stout), *tiurr* (high-water mark).
- <eo> is found in just five words: *deoc* (to suck), *deoch* (drink), *sgleog* (whack), *steorn* (to guide).

With regard to minimal pairs, it was found that, of the 237 S_B contexts that appear in the wordlist:

- There are 13 two-way contrasts between <ea> and <io>, e.g. *fleadh* (banquet) - *fliodh* (weed).
- There is just the one two-way contrast between <ea> and <eo>: *deach* (went) - *deoch* (drink).
- There are no minimal pairs involving a contrast between <io> and <iu>, between <io> and <eo>, or between <iu> and <ea>.

These observations suggest that there is a fundamental two-way, height-based short vowel contrast operating in S_B contexts - high <io>, and mid-low <ea>. By default, these vowels are realised at surface level as

unrounded front vowels /i/ and /ɛ/,¹¹ but the effect of the exceptional <iu> and <eo> orthography is to indicate lexically marked words which are pronounced with the rounded back vowels /ju/ and /ɔ/. As was the case with the B_B contexts discussed in the previous section, there is a tendency to prefer lower vowels before a broad coda – only 25% of broad codas are preceded by <io> or <iu>, compared to 73% by <ea>.

Turning now to words which have both a slender onset and a slender coda (i.e. S_S contexts), there are no fewer than *six* orthographically distinct vowel sequences: <ei>, <i>, <e>, <eai>, <eoi> and <iui>, typically pronounced as /e/, /i/, /e/, /a/, /ɔ/, and /u/ respectively. However, the first two occur much more frequently than the others, suggesting again that the latter are statistical outliers. To be precise:

- <e> is found in only two words: *teth* (hot), *leth* (half), both of which have a (generally unpronounced) *-th* coda.
- <eai> is found in just four words: *beairn* (gap - genitive), *beairt* (loom - genitive), *ceaird* (craft), *feairt* (attention - genitive), all of which have a coda consisting of R plus a dental.
- <eoi> is found in just two words: *sgleoig* (whack - genitive), *smeoirn* (arrow end).
- <iui> is found in just one word: *sriuit* (tirade - genitive).

The data from minimal pairs is as follows:

- There are six two-way contrasts between <ei> and <i>, e.g. *beinn* (mountain) - *binn* (verdict).
- There is one two-way contrast between <e> and <ei>: *leth* (half) - *leith* (nerve), although these actually appear to be homophones.
- There are no minimal pairs involving <eai>, <eoi> and <iui>.

Finally, with regard to morphological slenderisation patterns:

- All words in <io> slenderise to <i>, e.g. *lios* - *lis* (garden).
- 60% of words in <ea> slenderise to <ei> and 37% to <i>, e.g. *creag* - *creig* (cliff), *beag* - *big* (small).

¹¹ Although in many contexts (e.g. before velarised sonorants), <io> is realised as /ju/ or /ju:/ (Bauer 2011: 531), e.g. *iolair* (eagle), *ionnsaich* (to learn), *rionnag* (star), *fionn* (fair).

Just as was the case with the S_B contexts discussed above, this data suggests a fundamental two-way height-based short vowel contrast operating in S_S contexts – high <i>, and mid-low <ei>. Words with <e> are simply an orthographic variant for <ei>, and the marginal words with <eoi> and <iui> (and possibly <eai> as well) must be considered to be lexically marked exceptions, involving surface back vowels, rather than the default front vowels /e/ and /i/.

8 Summary of the Gaelic short vowel inventory

My conclusions about the morphophonemic inventory of short stressed vowels in Gaelic monosyllabic words are summarised in Table 3. Lexically marked alternatives are placed in parentheses, with highly marked, exceptional ones in double parentheses.

| | S_S | S_B | B_B | B_S |
|------|--------------------|----------------|-----|------|
| high | i ((iui)) | (io) ((iu)) | (u) | ui |
| mid | ei ((e)) | ea ((eo)) | o | (oi) |
| low | ((eai)) ((eoi)) | | a | ai |

Table 3 – The short stressed vowel inventory in modern Gaelic

There is a basic three-way height contrast, between high, mid and low vowels (with the latter two values neutralised after a slender onset), along with general tendencies to avoid mid-vowels before a slender coda and to prefer lower vowels before a broad coda. Phonological roundedness and front-ness are almost always predictable from the broad/slender value of the onset and coda. Vowels tend to be unrounded both after a slender onset and before a slender coda, and low vowels are always unrounded. Similarly, vowels tend to be front after a slender onset and back after a broad one. However, there are a few lexically marked exceptions, where back vowels appear after a slender onset – represented orthographically as <iu>, <iui>, <eo>, <eoi> and <eai>.

The morphophonology of Gaelic vowels

To sum up, the orthography suggests that there are just *three* basic short stressed vowel morphophonemes – high, mid and low – along with a few lexically marked rarities after slender onsets: high-back (<iu>, <iui>); and mid/low-back (<eai>, <eoi>, <eo>). The short vowel morphophonemes in Scottish Gaelic can be argued to constitute a fundamentally one-dimensional, vertical vowel system, a claim similar to one which has previously been made about the abstract phonology of certain Irish dialects (Skerrett 1967; Bliss 1972; Ó Siadhail 1989: 36; cf. Ó Maolalaigh 1997).

I conclude this summary with a brief discussion of the long vowels and diphthongs, in relation to the short vowel inventory. A preliminary attempt at categorising these vowels is presented in Table 4, again using the orthographic representations to denote the vowel morphophonemes.

| | S_S | S_B | B_B | B_S |
|-------------|-------------|----------------|---------|-----------|
| high | ì ((iù)) | ìò (iù) | ù ao | ùì aoi |
| mid | èì (eòì) | ia eu | ò ua | òì uai |
| low | ((è)) | (eò) ((èa)) | à | ài |

Table 4 – The long stressed vowel inventory in modern Gaelic

A first conclusion we can draw about the long vowels and diphthongs is that they appear to fit reasonably comfortably into the same morphophonemic paradigm as the short vowels, although the precise featural specification of the diphthongs <ao(i)> and <ua(i)> requires further investigation. Another thing that is clear is that with the long vowels, the general tendencies to avoid high vowels before a broad coda, and to avoid mid vowels elsewhere are much weaker than with the short vowels, as is the tendency to avoid back vowels after a slender onset.

Finally, the slenderisation patterns of the S_B long vowels and diphthongs are also interesting: (a) <ia> usually slenderises to <èì>, but occasionally to <eòì> or <i>; (b) <eu> slenderises to either <èì> or <eòì>, in approximately equal proportions, but never to <i>; (c) <eò> mainly

slenderises to <eòì> but occasionally to <iùì>; and (d) <ìò> always slenderises to <ì>.

9 Implications for Gaelic language planning

One thing that stands out in this analysis of the deep phonology of Scottish Gaelic short stressed vowels is the fact that the morphophonology of Gaelic vowels works very differently from that of the consonants. Gorrie's (2011) treatment of consonant morphophonology was couched in terms of a dedicated inventory of abstract morphophonological features (broad/slender, fortis/lenis), which are then interpreted in terms of more concrete phonological features such as [palatal] and [continuant]. However, the morphophonemic treatment of Gaelic vowels offers no motivation for a special abstract feature set. Rather the process of interpreting vowel morphophonemes appears to be more a matter of using the general constraints to flesh out the highly underspecified lexical representations. For example, the underlying lexical representation of the adjective *fionn* (fair) will simply specify the vowel as being [high], with the extra features [back], [round] and [long] being added during the transition from deep to surface phonology by means of generic phonological rules, giving us a transition from morphophonemics to phonemics as follows:

- $F_{[\text{fortis, slender}]} V_{[\text{high}]} N_{[\text{fortis, broad}]} \Rightarrow /fj u: N/$

Thus, it may well be more accurate to say that whereas the lexical phonology of Gaelic consonants is morphophonemic in nature, that of vowels is closer to being straightforwardly phonemic, at least in terms of distinctive features.

Another thing that this analysis suggests is that Gaelic orthography is much closer to being a 'consonantal' writing system, like that for Hebrew and Arabic, than is commonly realised, since the vowel graphemes appear to play just as much of a role in distinguishing broad and slender consonant morphophonemes as they do in distinguishing different vowels. Were the Gaelic language in a better state of sociolinguistic health, it would thus not be unreasonable to propose a complete neutralisation of vowel distinctions in the orthography, with just four grapheme sequences needed to represent the short stressed vowels - say <e>, <ea>, <a> and <ai>.¹² The relative lack of minimal pairs in Gaelic would allow for a fluent speaker of the language

¹² Though, of course, this would restrict access to more traditional forms of the language in older books.

to easily resolve the resulting ambiguities in real time from the context, although certain pairs, e.g. *lach* (a duck) versus *luch* (a mouse), might well cause problems. This would be a ‘writer-centric’ orthographic system, making it as easy as possible for writers to (learn to) encode speech in written form, whilst making it slightly more difficult for readers (especially learners) to decode the written language. A less extreme variant on this writer-centric approach would partially encode the height distinction by allowing *two* grapheme sequences for each syllable context, but neutralising the marginal cases – replacing <iu> with <io>, <oi> with <ui>, etc.

However, the Gaelic language is demonstrably *not* in a state of good sociolinguistic health. There are no monolingual speakers of Gaelic remaining, and Gaelic appears to be irrevocably disappearing as a community and family language in its traditional rural heartlands, where intergenerational transmission is fundamentally broken (Munro 2010). In the future, it is highly likely that Gaelic will be mainly spoken in urban areas, in speech communities formed by middle-class ‘nouveaux Gaels’ (Morgan 2000) who have learned the language either as adults by means of second language learning or as children through Gaelic-medium immersion schooling. From this perspective, the role of the orthographic system in language acquisition planning and language revitalisation needs to be considered seriously, giving due weight to the fact that in future most Gaelic-speakers will have learned to speak and read the language at the same time, and therefore that, to a large extent, they will have had to deduce the underlying morphophonology of the language from its written representation, rather than learning to read and write having already acquired the morphophonemics during infancy. Thus, the more cues the written form can give learner-readers about the morphophonemics the better, and the common pedagogical tendency to fixate on surface-phonological transparency should be resisted.¹³

Aside from the question of orthographic reform, there are at least three other ways in which better awareness of Gaelic vowel morphophonology can influence current and ongoing language planning initiatives for the language. From the perspective of acquisition planning, having a good statistical model of phonological markedness will be useful for determining the order in which to present vocabulary items in the very early stages of language acquisition, with less marked words being

¹³ For example, the modern practice of representing non-diphthongal surface compensatory lengthening explicitly in the orthography using the grave accent, e.g. *bòrd* (a table), *ceàrr* (wrong). Irish practice differs here, cf. the official standard *An Caighdeán Oifigiúil*.

introduced first, and the exceptions only being introduced once students have a firm grasp of the basics of Gaelic morphophonemics. Similarly, in terms of literacy planning, having a good account of the relative frequencies of different phoneme-grapheme vowel correspondences is a prerequisite for developing more effective phonics-style approaches to learning to read and write Gaelic in the initial years of Gaelic-medium education (cf. Stòrlann 2005).

Finally, this kind of work is also important for the development of speech and language technology for a lesser-resourced language like Gaelic. Creating these systems generally is a two-step process, involving the formulation of a rough model of the linguistic phenomenon in question, which is then slowly refined by machine learning algorithms which expose the model to real-life linguistic data and ‘train’ it to become more accurate. The problem for lesser-resourced languages is that there simply are not enough electronic corpora of text or speech available to perform the necessary training – hence the onus is on the people who design the initial model to make it as good as possible. With this in mind, it is expected that the model of vowel morphophonemics presented in this paper will prove useful in ongoing work to develop automatic text-to-speech and speech-to-text systems.

10 Conclusions

The starting point for this investigation was Gorrie’s (2011) independently motivated, feature-based analysis of the morphophonemics of Gaelic consonants, which turns out to be very close in many respects to what we would expect based solely on a careful analysis of the traditional Gaelic orthography. The central question was whether the traditional orthography has anything to teach us about the morphophonemics of Gaelic vowels – whether it might be possible to reverse-engineer the morphophonemic structure of Gaelic vowels from their traditional orthographic representation. Based on a gold-standard list of 2,264 Gaelic monosyllabic words derived from a combination of electronic and paper-based dictionaries, the short stressed vowel taxonomy presented in Table 3 was developed. This taxonomy is characterised by a fundamental three-way distinction between high, mid and low vowels (with the latter two values neutralised after a slender onset), but is flexible enough to allow for a couple of lexically marked rarities occurring after slender onsets – high-back and mid-back. This analysis was shown to have a number of

The morphophonology of Gaelic vowels

interesting implications for ongoing language planning efforts for Scottish Gaelic, in particular orthographic reform to support language acquisition.

References

- Bauer, Michael. 2011. *Blas na Gàidhlig*. Glasgow: Akerbeltz.
- Black, Ronald. 2010. 'Gaelic Orthography: The Drunk Man's Broad Road'. In Watson, Moray and Michelle Macleod (eds.) *The Edinburgh Companion to the Gaelic Language*. Edinburgh University Press, 229-61.
- Bliss, Alan J. 1972. 'Languages in contact: some problems of Hiberno-English'. *Proceedings of the Royal Irish Academy*, 72: 63-82.
- Bosch, Anna. 2010. 'Phonology in Modern Gaelic'. In Watson, Moray and Michelle Macleod (eds.) *The Edinburgh Companion to the Gaelic Language*. Edinburgh University Press, 262-82.
- Gorrie, Colin. 2011. 'Adjective agreement in Gaelic: A case for morphophonological features'. In Carnie, Andrew (ed.) *Formal approaches to Celtic linguistics*. Cambridge Scholars Publishing, 321-35.
- Halle, Morris and Alec Marantz. 1993. 'Distributed morphology and the pieces of inflection'. In Hale, Kenneth and S. J. Keyser (eds.) *The View from Building 20*. Cambridge MA: MIT Press, pp. 111-76.
- Mark, Colin. 2004. *The Gaelic-English Dictionary*. Routledge.
- MacAulay, Donald. 1968. 'Palatization of labials in Scottish Gaelic and some related problems in phonology'. *Scottish Gaelic Studies* 11: 72-84.
- Morgan, Peadar. 2000. 'The Gaelic is dead; Long live the Gaelic: The changing relationship between native and learner Gaelic users'. In McCoy, Gordon and Maolcholaim Scott (eds.) *Aithne na nGael: Gaelic Identities*. Belfast: Institute of Irish Studies / Ultach Trust, 126-33.
- Munro, Gillian. 2010. *Barail agus Comas Cànnain sa Choimhearsnachd: Siabost*. Bòrd na Gàidhlig / Sabhal Mór Ostaig.
- Ó Maolalaigh, Roibeard. 1997. *The Historical Short Vowel Phonology of Gaelic*. Unpublished PhD thesis, University of Edinburgh.
- Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical Structure and Dialectal Variation*. Cambridge University Press.
- Oftedal, Magne. 1965. 'On "palatalized" labials in Scottish Gaelic'. *Scottish Gaelic Studies* 10: 71-81.
- Skerrett, R. A. Q. 1967. 'Notes on the dialect of the Inishkea islanders'. *Studia Celtica* 2: 196-201.

The morphophonology of Gaelic vowels

- SQA. 2009. *Gaelic Orthographic Conventions*. Scottish Qualifications Agency, Glasgow.
- Stòrlann. 2005. *Facal agus Fuaim*. Stornoway: Stòrlann.
- Ternes, Elmar. 2006. *The Phonemic Analysis of Scottish Gaelic - Third edition*. Dublin: Institute for Advanced Studies.