

‘Risk-free’ corpus planning for Scottish Gaelic?

Collaborative development of basic grammatical norms for 21st century speakers

Roibeard Ó Maolalaigh & Mark McConville

1. Background

1.1. The challenge of corpus planning for Scottish Gaelic

The fundamental aim of corpus planning (at least as this term is used among Gaelic language planners in Scotland) is to ensure that the language is stable enough and expressive enough to support the needs and ambitions of its speakers. Ultimately, it supports, and is a prerequisite for, effective status planning. In the words of the first *National Plan for Gaelic*, it strengthens ‘the relevance and consistency’ of language (Bòrd na Gàidhlig 2007: 17). Despite many advances in Gaelic language development in Scotland since the 1970s, most notably Gaelic-medium education, there has been no co-ordinated consistent approach to Gaelic corpus planning – as first pointed out by McLeod (2004). Gaelic corpus development has lagged significantly behind other areas of language planning. Estimates place Bòrd na Gàidhlig’s expenditure on corpus planning initiatives at around 1% of its total annual budget (McConville, McLeod and Ó Maolalaigh 2011: 2). In contrast to other minority languages, including Irish Gaelic, Scottish Gaelic has never had a formal overarching institution or body with responsibility for the codification and elaboration of Gaelic.¹

This uncentralised approach can be traced back to the 17th century, when individuals and organisations began to work, largely independently, on major projects such as the translation of the Bible into Gaelic and the compilation of dictionaries and grammars.² These works became by default the pillars of codification over the past four centuries. Despite a lack of co-ordination, it is important to note that a great deal was nevertheless achieved during this period.

Many of the early pioneers in corpus development were alumni of the University of Glasgow, for example Dugald Campbell (1599–1673), Minister of Knapdale, who contributed to the translation of the first fifty metrical psalms (*An Ceud Chaogad*), published in Glasgow in 1659 by the Synod of Argyll, and one of the earliest texts in which vernacular features of Scottish Gaelic consistently emerge (Synod of Argyll 1659; Scott 1923: 15–16; Grimble [1983] 1987; Thomson 1976). The Rev. William Shaw (1749–1831), another alumnus of Glasgow, published a Gaelic grammar (Shaw 1778), with the encouragement of Samuel Johnson, followed shortly afterwards by his *Galic and English Dictionary* (Shaw 1780;

¹ However, the Scottish Certificate of Education Examination Board did establish a sub-committee of the Gaelic Panel in 1976 to investigate inconsistencies in modern Gaelic orthography. The guidance of the sub-committee, chaired by Donald MacAulay, was published in August 1981 under the title of *Gaelic Orthographic Conventions*, which is generally referred to as *GOC*. This was revised and republished by the Scottish Qualifications Authority in 2005 and 2009. Although incomplete in a variety of ways *GOC* has proved to be a valuable resource for schools, colleges and universities. For a review of *GOC* (2009), see Cox (2010).

² For a historical survey of the development of Scottish Gaelic orthography, see Black (2010).

MacDonald 1979; Cram 1996; Macleod 2017). Thus, the connection of the University of Glasgow with initiatives in corpus development has a long and distinguished history.

1.2. National Plan for Gaelic 2007–2012: the ‘Gaelic Language Academy’

One of the main tasks that Bòrd na Gàidhlig (Scotland’s statutory development agency for Gaelic language policy) has to undertake under the Gaelic Language (Scotland) Act 2005 is to prepare a national Gaelic development plan every five years, with the first one having been published in 2007. This document laid out no fewer than 14 ‘priority areas’ for Gaelic development, two of which were connected to corpus planning goals. The 12th priority area, ‘Gaelic orthographic, terminological and place-name development’, was justified in the following terms (Bòrd na Gàidhlig 2007: 34):

Gaelic already has a well-developed grammar and a writing system that is relatively clear. However, all languages evolve over time, and it is important that the grammar and writing system of Gaelic is further developed, standardized and disseminated.

It was associated with a ‘key project’ which pledged that the Bòrd would ‘investigate the most suitable structure for a Gaelic language academy in order to ensure the relevance and consistency of Gaelic, including place-names’ (Bòrd na Gàidhlig 2007: 35). The 14th priority area, ‘Survey and research’, listed four research priorities for corpus planning:

- ‘research the nature of contemporary Gaelic vocabulary and grammar to inform future developments’
- ‘research, agree and promote formal standards for Gaelic spelling, names, signs, grammar and official register’
- ‘research and develop an authoritative historical dictionary of Gaelic’
- ‘research, develop and promote a national gazeteer of Gaelic place-names’

This represented a highly ambitious agenda for corpus planning activity between 2007 and 2012. How successful were Bòrd na Gàidhlig at achieving these goals over this period?

A close reading of Bòrd na Gàidhlig’s official annual reports allows us to trace what progress was made in setting up a Gaelic language academy over these five years:

- Annual Report 2008–09 states, very concisely, that they had ‘progressed discussion on the establishment of a Gaelic Academy’ (p. 26).
- Annual Report 2009–10 (p. 34) contains a reference to ‘a virtual Gaelic Language Academy being investigated by Research Committee with a view to establishing an interim structure’ (p. 34), though there is no explanation given for what ‘virtual’ means in this context.
- Annual Report 2010–11 goes into a little more detail, and refers to consulting on ‘a possible structure for a Gaelic Language Academy, which the Bòrd expects to provide an authoritative home for Gaelic corpus development and a new opportunity for increased co-operation between projects’ (p. 18). The report goes on to note: ‘Progress

on this has been slower than expected and it is now anticipated that the public consultation will take place as part of the National Gaelic Language Plan 2012–17 consultations. It is hoped that the Gaelic Language Academy will be launched in June 2012.’

- Annual Report 2011–12 contains no explicit mention of a ‘Gaelic language academy’, resorting to generic vague labels such as ‘an authoritative structure to provide strategic guidance for Gaelic Corpus developments’.

When the second National Plan for Gaelic was published in 2012, there was again no explicit reference to a ‘Gaelic language academy’. Instead, it referred to ‘co-ordination of a range of language initiatives’ through what was referred to as ‘a Corpus Development Forum’.

In conclusion, over the course of the first five-year National Plan for Gaelic, at least from an external perspective, there does not appear to have been a huge amount of concrete progress made in developing or coordinating Gaelic corpus planning activities at a national level, identified as a key priority in the plan. Perhaps the most immediately noticeable development over the five years was one of what might usefully be referred to as ‘rhetorical deflation’ – morphing from ‘language academy’ through ‘virtual language academy’ to ‘authoritative structure’ to ‘corpus development forum’.

1.3. ‘Survey of Gaelic Corpus Technology’ (2010)

Further impetus for the establishment of some kind of Gaelic language academy was provided by the ‘Survey of Gaelic Corpus Technology’ project, commissioned by Bòrd na Gàidhlig in 2009, and carried out by a research team at the University of Glasgow (Bauer, Ó Maolalaigh and Wherrett 2010). The main purpose of this research was to survey the views of Gaelic language professionals about the need to develop speech and language technologies for Gaelic, such as machine translation and speech recognition. However, one of the main conclusions of this substantial report was that, before significant progress could be made with Gaelic corpus technology, there was a need to get Gaelic’s house in order with respect to a co-ordinated infrastructure for corpus development more generally:

The current focus of Gaelic SALT [Speech and Language Technology] is seriously mismatched with where it should be, focusing on dictionaries and word lists rather than a comprehensive and integrated approach to codification and standardisation, essential SALT, corpus and lexicographical tools.

[T]he structures currently in place are disparate and uncoordinated. Best practice shows that there ought to be a formal Gaelic Academy that owns the codification (orthography, grammar and terminology) and is final arbiter on matters of technical aspects relating to the formal language. (Bauer, Ó Maolalaigh and Wherrett 2009: 2)

1.4. ‘A Way Forward for Gaelic Corpus Planning and the Gaelic Language Academy’ (2011)

Perhaps in response to this, in 2011, Bòrd na Gàidhlig’s Language Academy Working Group tasked the Board of Celtic Studies (Scotland) with putting forward some ideas for discussion that might progress thinking on Gaelic corpus planning and what the proposed Gaelic Language Academy might look like. In response to this invitation, the authors wrote, in collaboration with Wilson McLeod, a White Paper entitled ‘A Way Forward for Gaelic

Corpus Planning and the Gaelic Language Academy’, which was presented to the Working Group in November 2011 (McConville, McLeod and Ó Maolalaigh 2011).

Three recommendations were made with initial suggestions for discussion on how to progress each:

1. Bòrd na Gàidhlig should commission a 6–12 month investigative survey into corpus planning for Gaelic, in order to: (a) establish an appropriate linguistic foundation for the work (i.e. a statement of basic principles); and (b) survey and evaluate the work that has already been done.
2. Subsequently, the Bòrd should establish an independent, eight-member Gaelic Language Academy to oversee and regulate the corpus planning process.
3. Simultaneously, the Bòrd should consider establishing a professional Gaelic language research institute to carry out corpus planning for Gaelic, under the supervision of the Academy.

This White Paper made some very concrete and bold proposals, which for the first time, brought into sharp focus the scale of resources that might be needed to support professional corpus planning. While these proposals were made towards the end of 2011, it may be no coincidence that the term ‘Gaelic language academy’ disappears from official discourse very shortly thereafter!

1.5. *Dlùth is Inneach*

Although Bòrd na Gàidhlig had begun to avoid the term ‘Gaelic language academy’, they nevertheless took on board the ideas presented in the 2011 White Paper and agreed that further research, discussion and debate were needed. In the summer of 2012, they commissioned a new one-year research project. They asked for an investigative survey to be carried out into, not only the appropriate linguistic foundations for Gaelic corpus planning, but also asking for recommendations on a suitable institutional framework as well.

The two main research questions were:

- What corpus planning principles are appropriate for Gaelic?
- What effective coordination would result in their implementation?

We were keen to adopt a collaborative approach to this national challenge, and proposed that the Universities of Glasgow and Edinburgh should collaborate and put forward a joint bid under the auspices of the inter-university project *Soillse* – the National Research Network for the Maintenance and Revitalisation of Gaelic Language and Culture. We were fortunate enough to win the tender to carry out this research project during 2013. This was carried out by the authors and Susan Bell (now Ross) at the University of Glasgow, and Wilson McLeod at the University of Edinburgh.

As this project was to get to the heart of the language and its future, we decided to brand the project as *Dlùth is Inneach*, i.e. the ‘warp and the woof / weft’ – a borrowed metaphor from the traditional domain of weaving. This metaphor comes from the celebrated Gaelic author, the Rev. Donald Lamont (1874–1958), who strongly recommended that writing in Gaelic should use *fìor dhlùth is inneach na Gàidhlig* ‘the true warp and weft of Gaelic’. (Lamont 1960: 166, 168)

In this research, we wanted to emphasise the importance of what Joshua Fishman refers to as ‘risk-free corpus planning’, i.e. that innovations in corpus development ‘should be undertaken slowly and carefully, with a good understanding of the speech community’s language (Fishman 1991: 351). We were keen to avoid what Fishman refers to as ‘corpus planning that hinders’ and therefore to avoid corpus planning that might actively accelerate language shift and potentially undermine progress made in other aspects of the language. Importantly, this would involve collaboration with non-academic organisations and particularly with the Gaelic speech community itself.

Thirty-nine ‘focused conversation’ events were conducted across Scotland (Glasgow, Inverness, Lewis, Harris, Uist and Skye), involving 184 participants. These conversations dealt with three areas:

- linguistic foundations for Gaelic corpus development
- corpus resources for Gaelic
- institutional foundations for Gaelic corpus development

Sixteen conclusions resulted from this research, some of the most significant of which are discussed in the following subsections (see Bell et al. 2014: Part C).

Linguistic foundations

The accepted model for ‘good’ Gaelic (both formal and informal) is held to be the traditional Gaelic of speakers born before the 1960s. The dominant ideology amongst Gaelic speakers of all ages and geographical areas can be seen to be a limited form of ‘retrophilia’, which we characterised as ‘retro-vernacular’, i.e. an attachment to the traditional form of the language used by highly fluent traditional speakers. Research by Brian Ó Curnáin in Ireland (Ó Curnáin 2007: 58–60; 2016) has shown empirically that traditional Irish Gaelic norms begin to be eroded in the speech of those born from the 1970s onwards; a similar development is evident in Scottish Gaelic.

The generation gap between these ‘model speakers’ and the younger English-dominant bilinguals is keenly felt and manifests itself in the erosion of traditional grammar, idiom and the lexicon in the latter cohort. Bridging this generation gap is seen as an urgent priority for corpus development and for securing the future strength and sustainability of the language.

Corpus resources

Gaelic users are very aware of the significant ‘resource deficit’ in the language, in particular the lack of:

- explicit guidance on detailed aspects of grammatical usage
- an online ‘one-stop-shop’ for Gaelic resources with authoritative trustworthy advice on lexical and grammatical usage
- greater consistency in new terminology, avoiding a range of competing synonyms created by different organisations.

Institutional foundations

It emerged from our research that a successful Gaelic corpus development framework should have three sources of legitimacy in order to secure majority buy-in for decisions made in relation to grammar, terminology and so on:

- popular legitimacy
- scientific legitimacy
- political legitimacy

This suggested three different groups of key stakeholders:

- community-recognised ‘model’ Gaelic speakers
- language scientists, such as grammarians, phoneticians, lexicographers, and sociolinguists
- language status agents, i.e. representatives of Gaelic stakeholder organisations such as Bòrd na Gàidhlig, BBC, Stòrlann, etc.

Taking all of this on board, we recommended the establishment of a ‘independent, participatory Gaelic corpus development framework which embodies the ideology of the Gaelic language community’. It should have representation and involvement from the three key stakeholder groups just mentioned – essentially a ‘consensus-driven partnership’ of the key stakeholders in Gaelic corpus development. We also recommended a two-year pilot that would essentially test the tripartite model we proposed.

The final 222-page report (Bell et al. 2014) was submitted in February 2014 and approved later that year by the Bòrd.

1.6. Comataidh Comhairleachaidh Cànan (CCC) and LEACAG (2015–18)

Bòrd na Gàidhlig took on board many of the recommendations of the *Dlùth is Inneach* project. In late 2015, they established a ‘corpus steering group’ (Buidheann Stiùiridh Corpais, or BSC), of seven ‘model speakers’ each of whom had significant experience of working in Gaelic stakeholder organisations, thus combining popular and political legitimacy. This body renamed itself the ‘language advisory committee’ (Comataidh Comhairleachaidh Cànan, or CCC) in 2017, and represents the latest metamorphosis in the evolution of the notion of a ‘Gaelic language academy’. In January 2016, the Bòrd issued another tender entitled ‘Gaelic Corpus Development (CR 15-12)’. The main outputs sought by the Bòrd were:

1. advice on Gaelic linguistics provided to a corpus steering group composed of accomplished Gaelic speakers
2. a description of the main grammatical issues faced by the modern language, in agreement with the corpus steering group
3. an online space for the coordination, evaluation and dissemination of new Gaelic terminology

The Bòrd was still very keen on a cross-institutional collaborative approach, and thus in early 2016, a Soillse consortium, led by the University of Glasgow but also including researchers from the University of Edinburgh and Sabhal Mòr Ostaig (University of the Highlands and

Islands), tendered successfully for this contract. We branded the project as ‘LEACAG’, a kind of acronym for *Leasachadh Corpais na Gàidhlig* (‘Gaelic Corpus Development’). The Gaelic word *leac* means ‘flagstone, slab, hearthstone’. LEACAG, with Gaelic diminutive *-ag*, was intended to signal a small stepping-stone – a pilot phase – on the longer-term journey of Gaelic corpus development.

2. Methodology of the LEACAG grammar project

As previously mentioned, as one of the central parts of the LEACAG project, we were given the task of identifying and investigating the ‘main areas of uncertainty’ in Gaelic grammar, and then of drafting some guidance that could be used in future by Gaelic speakers, especially those involved in education and broadcasting. Due to the complex nature of this task, and the fact that we would be working within a multi-institutional team, we decided at the very start to split this project into four, more or less independent sub-projects:

1. The first sub-project involved a survey of professional users of Gaelic, to determine the ‘main areas of uncertainty’ in Gaelic grammar. This work was undertaken by Will Lamb, Wilson McLeod and Charles Wilson from the University of Edinburgh, over the course of the first 10 months of the project.
2. This was then followed by the second sub-project, undertaken by Domhnall Uilleam Stiùbhart from Sabhal Mòr Ostaig (UHI), the Gaelic college on Skye. He conducted a survey of native Gaelic speakers living in Skye and the Western Isles, to determine their views and opinions on the ‘main areas of grammatical uncertainty’.
3. Simultaneous to this, the third sub-project was undertaken by Susan Ross, Mark McConville and Roibeard Ó Maolalaigh from the University of Glasgow. This entailed conducting a series of detailed corpus investigations into the same ‘main areas of uncertainty’.
4. Finally, the fourth sub-project involved bringing together all strands of evidence collected and to draft simple but authoritative grammatical guidance.

The methodology of each of these sub-projects is discussed in the following subsections.

2.1. Identifying the ‘main areas of uncertainty’

In autumn 2016, the Edinburgh team drafted an online questionnaire involving 24 constructions that they thought might be among the ‘main areas of uncertainty’ in Gaelic grammar. This questionnaire involved 137 linguistic examples in total, and respondents were asked to consider each one and classify it as either ‘acceptable’, ‘unacceptable’, ‘acceptable in some circumstances’, or ‘not sure’. Where respondents judged an example to be ‘acceptable in some circumstances’, they were invited to explain this in more detail. In addition, for each of the 24 potential issues, respondents were asked explicitly whether they considered this to be one of the ‘main areas of uncertainty’ in Gaelic grammar, and if so, why.

The link to this online questionnaire was sent to 94 professional users of Gaelic, and 27 full responses were received back. After the questionnaire results had been analysed, they were discussed in more detail in the course of interviews with senior experts with specialist knowledge of particular areas of Gaelic usage – mainly education, but also broadcasting,

publishing and translation. In order to determine which of the 24 potential ‘areas of uncertainty’ were the most important, a variety of different weighting schemes were tested out. The various rankings that resulted from these weightings were presented to CCC at a special meeting in January 2017, and they agreed a final list of eleven constructions that they thought most merited further investigation.

The eleven constructions were as follows:

1. genitive case
2. dative case
3. variation in irregular verb forms
4. inversion of direct objects
5. impersonal structures with *rach*
6. lenition of verbal nouns
7. forms and choice of prepositions
8. relative clauses with prepositions
9. forms with numbers and nouns
10. structures expressing ‘if’
11. direct object forms with *ga*

Two of these are discussed in detail below – the dative case (2) and impersonal structures with *rach* (5).

2.2. Survey of traditional Gaelic speakers

After the eleven ‘main areas of uncertainty’ had been finalised and approved, these were then handed over to the researchers at the other universities, so that they could investigate these constructions with reference to the usage of native speakers in the islands, and to modern Gaelic literary practices.

In summer and autumn 2017, the Skye team conducted a survey into the views and attitudes of native speakers of Gaelic, with respect to these ‘main areas of uncertainty’. Five consultation events were held in Skye, Lewis and the Uists. Each event involved the researcher giving an hour-long talk on a topic of local historical or cultural interest, at the end of which attendees were invited to fill out a questionnaire on Gaelic grammar, incentivised merely by cups of tea and local baking.

This questionnaire was an abridged version of that used by the Edinburgh team in the initial survey of professional users, reduced to just those eleven constructions which had been identified as meriting further study. As in the previous survey, respondents were asked whether they found the examples to be ‘acceptable’, ‘unacceptable’, ‘acceptable in some circumstances’, or ‘not sure’. The questionnaires were completed under strict but good-humoured exam conditions, with discussion being discouraged. The aim was to elicit near-instantaneous responses, avoiding too much grammatical self-reflection.

In all, 62 respondents completed the survey: 25 from Lewis; 15 from Skye; and 22 from the Uists. Of those stating how old they were, 14% were in their fifties, 40% in their sixties, and a further 23% in their seventies. Once the data had been collected, the research team conducted a descriptive statistical analysis of the responses, and produced a report on each of the grammatical constructions, which was submitted to CCC for approval.

2.3. *Corpas na Gàidhlig* investigations

However, as mentioned previously, we were not just interested in gathering the views of native speakers. We were also keen that the grammatical guidance be informed by the usage of modern Gaelic writers.

The University of Glasgow hosts the *Digital Archive of Scottish Gaelic (DASG)* website, a British Academy recognised and funded project. This archive was founded by Ó Maolalaigh in 2006 and was launched as an online website in 2014. One of the main components of *DASG* is *Corpas na Gàidhlig* – a fully searchable online corpus of 355 Gaelic texts, with a current total of around 28 million tokens (*Corpas na Gàidhlig*. 2014–19). The primary aim of the first stage of *Corpas na Gàidhlig* was to provide a comprehensive textual foundation for the long-term *Faclair na Gàidhlig* project, whose goal is to create a historical dictionary of Scottish Gaelic (Pike & Ó Maolalaigh 2013). However, a secondary aim of *Corpas na Gàidhlig* has always been to facilitate detailed linguistic research into all aspects of the language, and hence it was the best available resource for the LEACAG grammar project corpus investigations (Ó Maolalaigh 2016).

For a variety of reasons, it was decided not to use the whole of *Corpas na Gàidhlig* for this sub-project. Firstly, 28 million tokens was simply too much data to analyse in such a short period of time. And secondly, we were keen that our guidance should reflect the ‘retrovernacular’ language ideology identified in the course of the *Dlùth is Inneach* consultation project, as discussed previously. With this in mind, we restricted our attention to the sub-corpus of all the texts published after 1950. This sub-corpus consisted of 67 texts involving a total of around 2.5 million tokens. The vast majority of these texts were published from 1970 onwards – very much reflecting the renaissance in Gaelic publishing after the foundation of the Gaelic Books Council (Comhairle nan Leabhraichean) by Professor Derick Thomson at the University of Glasgow in 1968.

Using this sub-corpus, carefully defined search queries were presented to the online corpus engine, with the results being extracted and imported into spreadsheets. The search results were classified into different patterns of usage, and a descriptive statistical analysis was carried out for a range of hypotheses relating to the areas of uncertainty. Finally, a comprehensive report was written for each construction, and these reports were submitted to CCC in December 2017.

2.4. Drafting of provisional grammatical guidance

By this point in the LEACAG grammar project, we had three main ingredients:

1. a list of eleven ‘main areas of uncertainty’ in contemporary Gaelic grammar, identified after comprehensive consultation with professional users of Gaelic, and agreed with CCC

2. the results of the survey into the views and attitudes of native Gaelic speakers living in the Western Isles and Skye, with respect to these eleven constructions
3. the results of comprehensive corpus investigations, deriving from careful study of the (mainly formal) Gaelic used by writers since 1950

The aim of the final part of the project was to synthesise all of this evidence into draft grammatical guidance, comprehensive in scope, but simple enough to be accessible to Gaelic users who may not be professionally trained academic linguists. This guidance was drafted as a collaboration between all three parts of the research team, and with extensive feedback from CCC. The final draft was approved in April 2018, and will form the basis of the next stage of the Gaelic corpus development project. The grammatical guidance was published in March 2019 on the *DASG* website (<https://dasg.ac.uk/grammar>).

The next two sections of the present paper report on two case studies involving ‘areas of uncertainty’ in contemporary Gaelic grammar, and how they were resolved.

3. Case study: Dative case

In Gaelic, nouns and adjectives which are governed by a basic preposition appear in the dative (or ‘prepositional’) case (rather than the nominative or genitive case) (Ó Maolalaigh [1996] 2008: 67–69; Lamb 2001: 29–32). Grammar books specify two main ways in which dative case is marked as different from the nominative:

- The masculine singular definite article lenites a following dative noun, unlike in the nominative case where there is no lenition. For example, nominative ‘the table’ is *am bòrd* in Gaelic, whereas the dative ‘on the table’ is *air a’ bhòrd*, where the noun *bòrd* has been lenited to *bhòrd*.
- Feminine singular nouns and adjectives are generally slenderised (i.e. palatalised) in the dative case, but not in the nominative. Take the example *bròg mhòr* meaning ‘a big shoe’ – since this is nominative neither the noun *bròg* nor the adjective *mòr* is slenderised. However, in the dative construction ‘in a big shoe’, *ann am bròig mhòir*, both the noun and the adjective have been slenderised, signified in Gaelic orthography by the insertion of the letter *i* before the final consonant.

Marking the dative case after the masculine singular definite article using lenition is still accepted as the ‘norm’ in modern Gaelic, with ‘innovative’ forms like **air am bòrd* being generally derided. However, there is an awareness that marking the dative via slenderisation of feminine nouns and adjectives is something which is applied inconsistently in modern Gaelic usage. For many speakers and writers, the dative forms of feminine singular nouns and adjectives are identical to the nominative, and examples such as *ann am bròg mhòr* (‘in a big shoe’), without slenderisation, occur commonly for the more traditional form *ann am bròig mhòir*, with slenderisation.

For example, Ó Maolalaigh ([1996] 2008: 67, 68) notes that slenderisation of feminine nouns and adjectives after prepositions ‘is disappearing in Gaelic and in many dialects is confined to a handful of [...] nouns and then usually only when the definite article precedes’, and that ‘slenderisation is far more common in literary Gaelic’. On the other hand, others have noted that contemporary published material continues to apply (orthographic) slenderisation, including in books aimed at children.

3.1. Dative case – Survey of professional speakers

With all this in mind, the questionnaire for the professional Gaelic speakers included a question about slenderised dative case forms. Respondents were presented with a series of prepositional phrases – some with slenderised feminine forms, and some without, some with the definite article, some without – and asked to what extent they considered them to be ‘acceptable’ or ‘unacceptable’:

- [indefinite, no slenderisation] *ann am bròg mhòr*
- [indefinite, slenderisation] *ann am bròig mhòir*
- [definite, no slenderisation] *anns a’ bhròg mhòr*
- [definite, slenderisation] *anns a’ bhròig mhòir*

All examples, both slenderised and non-slenderised, were considered as being acceptable to a clear majority of professional respondents, with the exception of the non-slenderised definites which half of the respondents judged unacceptable. In general:

- Non-slenderised forms are regarded as being *slightly* preferable to slenderised forms in *indefinite* contexts.
- But, slenderised forms are regarded as being *strongly* preferable to non-slenderised forms in *definite* contexts.

Slenderisation of dative feminine nouns and adjectives elicited relatively high scores in relation to all the various measures of ‘uncertainty’, and hence this construction was selected by CCC as necessitating further investigation, before any definitive guidance could be drafted.

3.2. Dative case – Survey of native speakers

As part of the survey of native Gaelic speakers, respondents were again presented with the same set of slenderised and non-slenderised feminine dative forms and asked to judge them for acceptability or unacceptability. Once the data had been analysed, it became clear that native Gaelic speakers are significantly less ‘conservative’ than the Gaelic language professionals, when it comes to slenderisation of dative feminine nouns and adjectives:

- In *indefinite* contexts, non-slenderised forms are *strongly* preferred by speakers.
- But in *definite* contexts, slenderised forms are *slightly* preferred.

However, the native speakers did agree with the professional speakers that the status of dative case marking is particularly unclear in contemporary Gaelic, with this phenomenon eliciting the highest ‘uncertainty rate’ in the whole survey.

3.3. Dative case – Corpus research

The question of slenderisation of dative feminine nouns and adjectives was also investigated with reference to the post-1950 sub-corpus of *Corpas na Gàidhlig*. As might be expected, recent Gaelic writers appear to be more ‘conservative’ than contemporary native Gaelic

speakers, in that the results of the corpus research were more closely aligned with the intuitions of the professional speakers:

- In *indefinite* contexts, non-slenderised forms are *slightly* preferred by writers.
- But in *definite* contexts, slenderised forms are *strongly* preferred, especially in the absence of an adjective.

3.4. Dative case – Drafting of grammatical guidance

It is clear from our research that usage with respect to slenderisation of dative feminine singular nouns and adjectives is involved in a historical process of language change from the old norm of obligatory slenderisation in all contexts to a future norm of non-slenderisation in all contexts. Written usage tends slightly more towards the conservative end of the spectrum, and hence these are the forms regarded as more acceptable by Gaelic professionals. On the other hand, native speakers tend towards the more progressive end, with slenderised forms seen as being ‘marked’ in some way. The key question for CCC was thus to decide which point in this continuum should be regarded as the ‘recommended basic norm’ for Gaelic learners and teachers, and how much flexibility should be allowed in different contexts.

Single-register approaches

In this particular case, as a pilot, we presented CCC with a full menu of options to choose from. At the very start of the LEACAG grammar project, CCC had specified clearly that they expected the grammatical guidance to be based around a core basic register of usage. In other words, they were keen to avoid guidance which recommended different forms in different contexts, or which used vague terms like ‘formal’ or ‘informal’. Thus, we started out by proposing three ‘single-register’ approaches. The first two of these were as simple as possible, in that they involved just accepting one of the two extremes and rejecting the other:

- In the ‘full slenderisation’ approach, dative feminine singular nouns and adjectives are *always* slenderised.
- Whereas in the ‘no slenderisation’ approach, dative feminine singular nouns and adjectives are *never* slenderised.

We also proposed a third, slightly more complicated single-register approach which aligned more closely with the evidence we had gathered from native speakers and from the corpus studies:

- In this ‘some slenderisation’ approach, dative feminine singular nouns and adjectives are slenderised in *definite* usages but not in indefinite ones.

At this point, the general consensus of CCC was that the ‘some slenderisation’ approach was closest to what they wanted. However, they thought that it needed elaboration, essentially since they did not want to be officially proscribing widely used forms. Thus, in spite of their previous stated desire for a single core basic register, they asked us then to propose a range of more subtle, ‘multi-register’ approaches for them to consider.

Multi-register approaches

Our first multi-register approach took ‘full slenderisation’ as the norm but permitted some variation in low-register contexts:

- In this ‘permissive full slenderisation’ approach, slenderisation of dative feminine singular nouns and adjectives is *recommended*, though non-slenderisation is to be *tolerated* in informal, especially indefinite, contexts.

Our second multi-register approach was the polar opposite of this – ‘no slenderisation’ is the norm but some variation can be ‘aspired to’ in high-register contexts:

- In this ‘ambitious no slenderisation’ approach, slenderisation of dative feminine singular nouns and adjectives is *not* officially recommended, though its use can be *encouraged* in formal, especially definite contexts.

Our third and fourth multi-register approaches took the intermediate ‘some slenderisation’ approach as the norm and proposed different directions of flexibility:

- In the ‘permissive some slenderisation’ approach, slenderisation of dative feminine singular nouns and adjectives is recommended in *definite* contexts but not in indefinite ones. However, non-slenderised definite forms are to be *tolerated* in informal contexts.
- In contrast, in the ‘ambitious some slenderisation’ approach, slenderisation of dative feminine singular nouns and adjectives is again to be *recommended* in definite contexts, but not in indefinite ones. In this case, however, use of slenderised indefinite forms is to be *encouraged* in more formal contexts.

To be wholly consistent, we also proposed an even more complicated, three-register approach, which distinguished both formal and informal alternatives to the basic unmarked norm:

- In the ‘fully flexible some slenderisation’ approach, slenderisation of dative feminine singular nouns and adjectives is again recommended in *definite* contexts but not in indefinite ones. However, non-slenderised definite forms are to be *tolerated* in informal contexts and slenderised indefinite forms are to be *encouraged* in formal ones.

When push came to shove, this last three-register approach was the one that CCC felt most comfortable supporting. Despite initially requesting a simple single-register approach as a matter of principle, in the face of actual sociolinguistic data, a complex approach involving variation in both formal and informal situations turned out to be the most acceptable solution.

4. Case study: Impersonal constructions with the verb *rach*

Scottish and Manx Gaelic (but not Irish) have an impersonal (i.e. passive-like) construction consisting of a form of the verb *rach* ‘go’ followed by a non-finite clause, containing an object followed by the inversion marker *a* (which lenites) and a verbal noun. Take, for example, the following simple Gaelic sentence:

- (1) Thog e an taigh.

build-PAST he the-MASC.SG house-(MASC).SG

‘He built the house’ (lit. ‘the house went to build(ing)’)

When the *rach*-impersonal construction is applied to this sentence, we get the following:

- (2) Chaidh an taigh a thogail.

go-PAST the-MASC.SG house-(MASC).SG PRT build-VN.LEN

‘The house was built’

The inversion marker *a* derives from the preposition *do* ‘to, for’. It signals a marked OV word order in what is generally a VSO language. Note that *chaidh* is the past independent form of the verb *rach* ‘go’, and that *t(h)ogail* is the verbal noun derived from the verb *tog* ‘build’.

However, Gaelic has other particles with the form *a*, which can sometime cause confusion. One of these is the third person possessive pronoun: *a* (‘his’), which lenites a following noun, and *a* (‘her’), which does not lenite a following noun but prefixes *h-* to vowels; the third person plural possessive pronoun is *an / am* (‘their’):

- (3) a **ch**as ‘his foot’ [ə xas]

a cas ‘her foot’ [ə k^has]

an cas ‘their foot’ [ə(ŋ) g^(h)as], am bas [ə**m** bas] ‘their palm’

One aspect of this confusion is that Gaelic speakers occasionally re-analyse the inversion marker *a* as a possessive pronoun in morphosyntactic agreement with the inverted object, resulting in hypercorrections like the following (note that *lùchairt* ‘palace’ is a feminine noun in Gaelic):

- (4) *Chaidh an lùchairt a **t**ogail.

go-PAST the-FEM.SG palace-(FEM).SG 3.SG.FEM build-VN

‘The palace was built’

- (5) *Chaidh na taighean **an** togail.

go-PAST the-PL house-PL 3.PL build-VN

‘The houses were built’

Note that the standard, non-hypercorrect usages would be the following:

- (6) Chaidh an lùchairt a **th**ogail.

go-PAST the-FEM.SG palace-(FEM).SG PRT build-VN.LEN

‘The palace was built’

- (7) Chaidh na taighean a **th**ogail.

go-PAST the-PL house-PL PRT build-VN.LEN

‘The houses were built’

Amongst the professional users we surveyed, there was a degree of uncertainty about the status of the *a* particle before the verbal noun in this impersonal construction. For example,

25% of users found the following non-traditional usage with the feminine noun *obair* ‘work’ to be acceptable:

- (8) Chaidh an obair a **dèanamh**.

go-PAST the-FEM.SG work-(FEM).SG 3.SG.FEM do-VN

‘The work was done’

In addition, 37% of users reported that the example with the plural noun *taighean* ‘houses’ in (5) is acceptable.

These results were mirrored to an extent by the traditional speakers. Notably, 42% of respondents reported that *chaidh na taighean an togail* was either acceptable or acceptable ‘in some circumstances’.

The results of the corpus investigations raised a few more intriguing questions. Three different versions of the impersonal construction were identified and analysed:

- (9) [Normal] Chaidh an taigh a **thogail**.

go-PAST the-MASC.SG house-(MASC).SG PRT build-VN.LEN

‘The house was built’

- (10) [Relative] an taigh a chaidh a **thogail**

the-MASC.SG house-(MASC).SG REL go-PAST PRT build-VN.LEN

‘The house that was built’

- (11) [Pronoun] Chaidh mo **thogail**.

go-PAST 1-SG build-VN.LEN

‘I was raised’

For the normal type, 98.7% of the examples were of the conforming type. There were twelve non-conforming examples found, nine of which involved plural nouns, for example:

- (12) Chaidh leth(-)bhreacan den Aithisg **an cuir**

go-PAST copy-PL of-the report-(FEM).SG 3.PL put-VN

‘copies of the report were put’

However, when used in relative clauses, the frequency of non-conforming examples increased to almost 5%, for example:

- (13) tro làimh a chaidh a **fliuchadh**

through hand-(FEM).SG REL go-PAST 3.SG.FEM wetten-VN

‘through a hand that was made wet’

- (14) bha uighean a chaidh **am breith**

be-PAST egg-PL REL go-PAST 3.PL lay-VN

‘there were eggs that were laid’

This suggests that the non-conformity may have originated in relativised examples, where the distance between the object noun and the verbal noun required the syntactic relationship to be strengthened or re-enforced – in this case via agreement in gender and / or number.

Finally, the corpus investigations threw up a number of non-conforming examples involving pronominal objects, where the object is realised as a personal pronoun rather than as a possessive. This is particularly the case with emphatic or otherwise modified pronominal objects:

(15) Chaidh mise a **thogail**³

go-PAST me-EMPH (PRT) raise-VN.LEN

‘I (emphatic) was raised’

(16) Chaidh iad seo/sin a **thogail**.

go-PAST them this/that PRT raise-VN.LEN

‘These/those were raised’

(17) Chaidh iad uile a **thogail**.

go-PAST they all PRT raise-VN.LEN

‘They were all raised’

(18) Chaidh e fhèin a **thogail**.

go-PAST he self PRT raise-VN.LEN

‘He himself was raised’

These patterns, though well known instinctively to fluent speakers, are not described explicitly in modern Gaelic grammar books or textbooks. Our corpus research allowed us to shed significant new light on the use of pronouns in the impersonal periphrastic construction in modern Scottish Gaelic, which will in turn enable the enhancement of existing learning and guidance materials in an area of confusion for both learners and native speakers alike. Overall, our research enabled us to provide CCC with clear and unambiguous guidance – in this case, guidance which was original and new to the language.

5. Next Steps for LEACAG

In this paper we have demonstrated how mutually beneficial collaborative work between universities and non-academic stakeholders has begun to have an impact on the linguistic and institutional foundations of corpus development for Scottish Gaelic. This approach is helping

³ The *a* is elided in speech and often not written following or preceding a vowel. It is retained here for clarity.

the language to keep up with the demands placed upon it by status planning initiatives in areas such as education and broadcasting.

The *Dlùth is Inneach* consultation project demonstrated the clear demand for corpus development among Gaelic speakers in Scotland – and in particular among the younger generations, who feel they are struggling to attain fully balanced bilingual language abilities. This work motivated Bòrd na Gàidhlig to establish a pilot Gaelic Language Academy, incorporating Comataidh Comhairleachaidh Cànanain (CCC) and the LEACAG support project. For two years, we worked in partnership to develop a range of corpus resources, in particular some draft provisional guidelines for recognised ‘areas of uncertainty’ in Gaelic grammar, such as those discussed in this paper – the dative case, and the use of impersonal constructions. These draft guidelines were delivered to the Bòrd in April 2018 and published in March 2019 on the *DASG* website (<https://dasg.ac.uk/grammar>). We are currently engaged in considering what should happen next, and how best to turn the pilot project into a more permanent corpus development framework.

Einar Haugen’s standard model for language planning recognises four stages of activity – selection, codification, implementation and elaboration. In the course of the LEACAG project, we have undertaken the first two stages – having identified some problems and challenges that need to be resolved, we *selected* (evidence-based) solutions for these problems, and *codified* these solutions in the form of grammatical guidance. The next stage of the project will necessarily involve both *implementation* and *elaboration*:

- How can we communicate our work to the wider Gaelic community, in particular to teachers and broadcasters, asking for feedback, and encouraging them to adopt these new norms?
- How can we develop the draft guidance from a short report on eleven ‘main areas of uncertainty’ into a comprehensive grammar of the modern Scottish Gaelic language – a clear desideratum in Gaelic Studies?

In all of our work, right from the very start of the *Dlùth is Inneach* project, we have been strongly motivated by what the pioneering sociolinguist Joshua Fishman has termed ‘risk-free’ corpus planning – corpus planning that ‘helps’ the revitalisation of a minority language rather than corpus planning that ‘hinders’ it (Fishman 1991: 351). We have proposed that placing equal weight on popular, scientific and political legitimacy in the development of corpus resources – the tripartite model – offers the best chance of doing ‘risk-free’ corpus development. Over the coming few years, we will discover how successful this approach has been.

References

- Baldauf, Richard B. 1989. 'Language Planning; Corpus Planning', *Annual Review of Applied Linguistics* 10: 3–12.
- Bauer, Michael, Ó Maolalaigh, Roibeard and Wherrett, Rob. 2009. *Survey of Gaelic Corpus Technology* (Inverness: Bòrd na Gàidhlig).
- Bell, Susan, McConville, Mark, McLeod, Wilson and Ó Maolalaigh, Roibeard. 2014. *Dlùth is Inneach: Linguistic Institutional Foundations for Gaelic Corpus Planning* (Inverness: Bòrd na Gàidhlig).
- Black, Ronald. 2010. 'Gaelic Orthography: The Drunk Man's Broad Road', in *The Edinburgh Companion to the Gaelic Language*, ed. by Moray Watson and Michelle Macleod (Edinburgh: Edinburgh University Press), 229–61
- Bòrd na Gàidhlig. 2007. *National Plan for Gaelic 2007–2012*. (Inverness: Bòrd na Gàidhlig).
- Bòrd na Gàidhlig. 2009. *Annual Report 2008–09*. (Inverness: Bòrd na Gàidhlig).
- Bòrd na Gàidhlig. 2010. *Annual Report 2009–10*. (Inverness: Bòrd na Gàidhlig).
- Bòrd na Gàidhlig. 2011. *Annual Report 2010–11*. (Inverness: Bòrd na Gàidhlig).
- Bòrd na Gàidhlig. 2012. *Annual Report 2011–12*. (Inverness: Bòrd na Gàidhlig).
- Bòrd na Gàidhlig. 2007. *National Plan for Gaelic 2012–2017*. (Inverness: Bòrd na Gàidhlig).
- Campbell, Dugald. 1659. *An Ceud Chaogad do Shalmaibh Dhaibhidh*. (Glasgow).
- Corpas na Gàidhlig, Digital Archive of Scottish Gaelic*. Available at <https://dasg.ac.uk/corpus>.
- Cox, Richard A. V. 2010. Review of *Gaelic Orthographic Conventions* [2009], *Scottish Gaelic Studies* 26, 162–73.
- Cram, David. 1996. 'William Shaw's Analysis of the Gaelic Language (1778): Text and Context', in *Linguists and their Diversions: A Festschrift for R. H. Robins on his 75th Birthday*, ed. by Vivien Law and Werner Hülsen (Münster: Nodus), pp. 245–74.
- DASG: Digital Archive of Scottish Gaelic*. Available at <https://dasg.ac.uk>.
- Fishman, Joshua. 1991. *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. (Clevedon: Multilingual Matters 76).
- GOC: Gaelic Orthographic Conventions* (Dalkeith: Scottish Certificate of Education Examination Board, 1981). Revised by the Scottish Qualifications Authority, 2005, 2009.
- Grimble, Ian. [1983] 1987. 'Campbell, Dugald', in *The Companion to Gaelic Scotland*, repr., ed. by Derick S. Thomson (Oxford: Basil Blackwell), p. 32.
- Haugen, Einar. 1983. 'The Implementation of Corpus Planning: Theory and Practice', in *Progress in Language Planning: International Perspectives*, ed. by Juan Cobarrubias and Joshua A. Fishman (Berlin: Mouton), pp. 269–89.

- Lamb, William. 2001. *Scottish Gaelic* (München: LINCOM Europa).
- Lamont, Donald. 1960. *Prose Writings of Donald Lamont 1874–1958*, ed. by Thomas M. Murchison (Edinburgh: Oliver and Boyd).
- MacDonald, Kenneth D. 1979. ‘The Rev. William Shaw – Pioneer Gaelic Lexicographer’, *Transactions of the Gaelic Society of Inverness* 50, 1–20.
- Macleod, Iseabail. 2017. ‘William Shaw, Controversial Gaelic Lexicographer’, in *Before the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster Triennial Meeting, Ayr 2015*, ed. by Robert McColl Millar (Aberdeen: University of Aberdeen), 81–97. Available at <https://www.abdn.ac.uk/pfrlsu/volumes/vol5-before-the-storm/>
- McConville, Mark, McLeod, Wilson and Ó Maolalaigh, Roibeard. 2011. *A Way Forward for Gaelic Corpus Planning and the Gaelic Language Academy*. Board of Celtic Studies (Scotland).
- McLeod, Wilson. 2004. ‘The Challenge of Corpus Planning in Gaelic Development’. *Scottish Language* 23, 68–92.
- Ó Curnáin, Brian. 2007. *The Irish of Iorras Aithneach, County Galway, Volume I*. (Dublin: Dublin Institute for Advanced Studies).
- Ó Curnáin, Brian. 2016. ‘Cróineolaíocht na Gaeilge Iarthraidisiúnta i gConamara, 1950–2004’, *Éigse* 39, 1–43.
- Ó Maolalaigh, Roibeard. [1996] 2008. *Scottish Gaelic in Twelve Weeks* (Edinburgh: Birlinn). Revised edition of *Scottish Gaelic in Three Months* ([Woodbridge]: Hugo).
- Ó Maolalaigh, Roibeard. 2016. ‘DASG: Digital Archive of Scottish Gaelic / Dachaigh airson Stòras na Gàidhlig’, *Scottish Gaelic Studies* 30, 242–62.
- Pike, Lorna and Ó Maolalaigh, Roibeard. 2013. ‘Faclair na Gàidhlig and Corpas na Gàidhlig: New Approaches Make Sense’, in *After the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster Triennial Meeting, Aberdeen 2012, Aberdeen, 19-20 July 2012*, ed. by Robert McColl Millar (Aberdeen: University of Aberdeen), 299–337. Available at https://www.abdn.ac.uk/pfrlsu/documents/Pike%20O%20Maolalaigh_2.pdf
- Scott, Hew. 1923. *Fasti Ecclesiae Scoticanæ*, iv (Edinburgh: Oliver and Boyd).
- Shaw, William (1778). *An Analysis of the Galic Language*, 2nd edn (Edinburgh: W. and T. Ruddiman).
- Shaw, William (1780). *A Galic and English Dictionary* (London: W. and A. Strahan).
- Synod of Argyll. 1659. *An Ceud Chaogad do Shalmaibh Dhaibhidh* (Glasgow: Aindra Aindrason).
- Thomson, R. L. 1976. ‘The Language of the *Caogad* (1659)’, *Scottish Gaelic Studies* 12.2, 143–82.