Faclair na Gàidhlig and Corpas na Gàidhlig: New Approaches Make Sense

Lorna Pike and Roibeard Ó Maolalaigh

1 Introduction

For minority languages in the twenty-first century increasingly overshadowed by their global counterparts, language maintenance and revitalisation are of paramount importance. Closely linked to these issues is the question of corpus planning. This essay will focus on two projects in Scottish Gaelic which will play a major part in preserving and maintaining the language by providing it with high quality lexicographical and research resources: Faclair na Gàidhlig and Corpas na Gàidhlig respectively; the essay concludes with a brief case study on Gaelic numerals which illustrates how Corpas na Gàidhlig can powerfully enhance our understanding of Gaelic. Faclair na Gàidhlig will be a comprehensive dictionary of Scottish Gaelic compiled on historical principles and with a structure similar to the *Oxford English Dictionary* (*OED*) and the *Dictionary of the Older Scottish Tongue* (*DOST*). The project was formally established in 2003 by an inter-university partnership comprising the universities of Aberdeen, Edinburgh, Glasgow, Strathclyde and the University of the Highlands and Islands through Sabhal Mòr Ostaig. It has received funding from a number of sources: Bòrd na Gàidhlig, the Carnegie Trust for the Universities of Scotland, the Gaelic Language Promotion Trust, the Leverhulme Trust, the Scottish Funding Council and the Scottish Government. Corpas na Gàidhlig, initiated in 2008, is the digital corpus on which the dictionary will be based and includes material from a variety of genres and periods of the language. It is being compiled at Glasgow University and is a constituent part of the larger Digital Archive of Scottish Gaelic / Dàta airson Stòras na Gàidhlig (DASG) project, directed by Professor Roibeard Ó Maolalaigh and established in 2006. DASG / Corpas na Gàidhlig is a British Academy recognised project, also supported financially by the University of Glasgow and Faclair na

Pike, Lorna and Roibeard Ó Maolalaigh. 2013. 'Faclair na Gàidhlig and Corpas na Gàidhlig: New Approaches Make Sense' In Cruickshank, Janet and Robert McColl Millar (eds.) 2013. *After the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting, Aberdeen 2012*. Aberdeen: Forum for Research on the Languages of Scotland and Ireland, 299-337. ISBN: 978-0-9566549-3-9

Gàidhlig; Comunn na Gàidhlig and the Gaelic Language Promotion Trust have contributed to the early stages. DASG / Corpas na Gàidhlig will provide a continually growing fully searchable database of Scottish Gaelic which will stimulate and enhance scholarly research on a scale not possible hitherto. Faclair na Gàidhlig and Corpas na Gàidhlig will be linked, complementary resources each enhancing the usability of the other. Dictionary users can do further searches on the corpus to access more evidence on the aspects of language in which they have an interest and corpus users will have easy access to a linked interpretive tool. In the wider context, they will also provide the basis for the development of technological, reference and educational resources, all of which are essential to future corpus planning for Gaelic.

1.1 Lexicographical context

Scotland possesses a world-class tradition in historical lexicography; see, for instance, Dareau (2012) and Macleod (2012). It has its roots in the nineteenth century in the work of Sir James Murray, from Denholm in the Scottish Borders, on the *OED*. In the early twentieth century, his colleague Sir William Craigie, from Dundee, transferred these skills to Scots. Craigie was the first editor of the *DOST*, which covers the Scots language from the twelfth century to 1700. The *DOST* completed publication to justified worldwide acclaim in 2002, and, along with its counterpart for the modern period, the *Scottish National Dictionary* (*SND*), it forms the authoritative foundation for the lexicography of Scots. Both the *DOST* and the *SND* have been digitised and are freely available online as the *Dictionary of the Scots Language* (*DSL*), now stewarded and updated by Scottish Language Dictionaries Ltd. Thus, in the early twenty-first century, the Scots language has globally accessible historical lexicographical resources which provide a foundation for the maintenance and revitalisation of the language.

However, Gaelic lexicography has fallen behind that of Scots and English despite the fact that it was flourishing in the nineteenth century when there was a prolific output (see Macdonald [1987] 1994; Pike 2008; Gillies and Pike 2012): two large dictionaries, one by Robert Armstrong (1825), the other by the Highland Society (1828) (the latter including a Latin-Gaelic section in addition to Gaelic-English and English-Gaelic sections); MacLeod and Dewar's

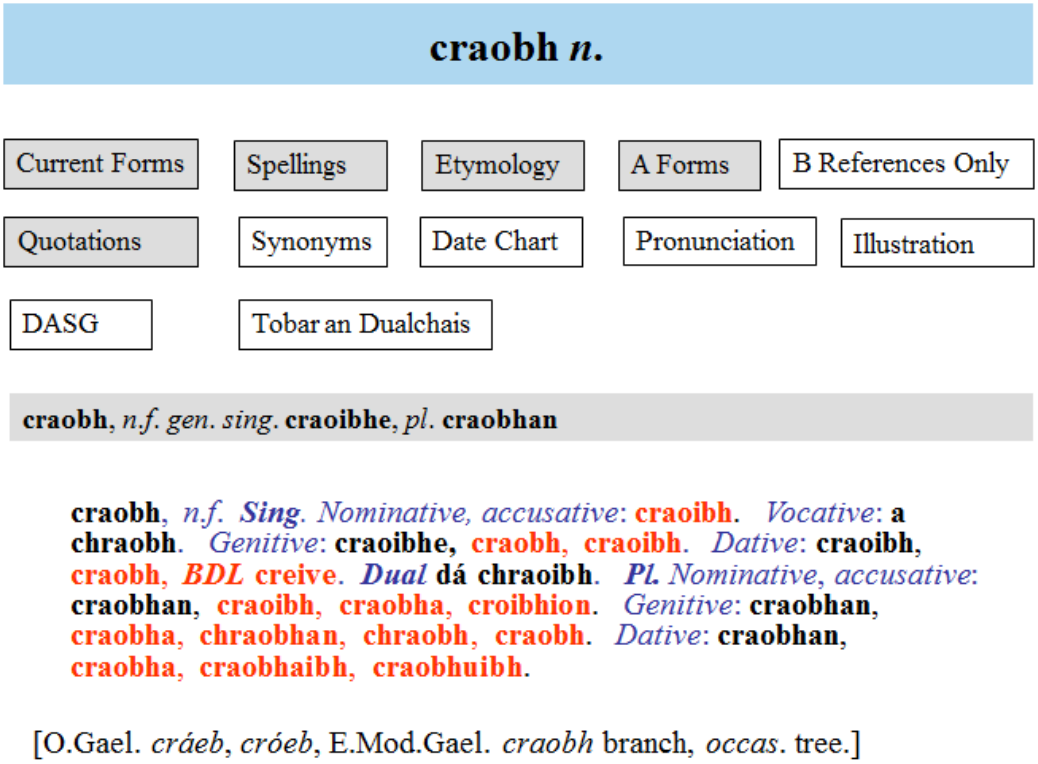
(1831) one-volume distillation of these dictionaries; MacAlpine's pronouncing dictionary (1832) (which predates the International Phonetic Alphabet); and Alexander MacBain's ground-breaking and scholarly etymological dictionary (1896) which brought Gaelic lexicography into the Indo-European context. The early twentieth century saw the publication of Edward Dwelly's acclaimed dictionary but, subsequently, Gaelic continued to be served only by smaller works compiled by individuals for a variety of reasons. This was in sharp contrast to the situation in English, Scots, Irish and Welsh where major historical dictionary projects were underway. The lack of an historical dictionary is arguably the greatest gap in Gaelic language resources today and it is essential to the survival of the language that this situation be remedied.

2 Faclair na Gàidhlig

The next priority in Scottish historical lexicography is to produce a dictionary for Gaelic, comparable to the *DOST* and the *OED*, to enable full understanding of the linguistic and cultural history of Scotland as a nation, as well as study of the interfaces between Gaelic and Irish, Scots, English and the Scandinavian languages. The Faclair na Gàidhlig project aims to create this resource. The challenge is immense but there is no doubt that this dictionary is a necessity; an undertaking essential to the sustainability and development of the Gaelic language. Faclair na Gàidhlig will be compiled in the electronic age which will open up new possibilities for both its compilation and its presentation. Its global accessibility will bring it to a much larger user group than previous historical dictionaries originally published in paper format. This stimulating prospect will provide many challenges for its lexicographers who will have to compile it to the highest lexicographical standards for the academic user but also make it easy to use for the lay person who chooses to dip into it on a regular or irregular basis.

The user-interface for the dictionary will form part of later project development but it is envisaged that users will be able to customise what is displayed on screen according to their needs. Full search facilities will lead them to onscreen options at the top of the entry for the word for which they have searched. The illustrations, excerpts from the sample noun entry for *craobh* ('a tree'), demonstrate the fields of information available to users; the

shaded boxes indicate those buttons which have been 'pressed' to give this display. The entry will begin with the current forms, e.g. the forms necessary to conjugate a noun, namely the nominative singular, the genitive singular and the nominative plural forms; or in the case of verbs, the root form and the verbal noun. This level of information will be of greatest use to learners of the language. This will be followed by a list of all the evidenced spellings with the obsolete forms indicated as such (shown in red in the illustration); and the



etymology displaying comparative material from other languages.

Illustration 1 – Current forms, spellings and etymology

The entry will then divide into two mains sections: section A Forms and section B Senses. Section A will include the earliest occurrence of each evidenced spelling and may include other examples if they are thought to be interesting from an orthographic point of view. Material will be included in this section for orthographic reasons only and will be particularly useful to those involved in formulating spelling policy. Forms will be sub-divided into case, number, tense, etc, and will be evidenced by fully referenced illustrative quotations.

craobh *n.*

A. Forms

Singular

***Nominative*: 13th-16th c** Madh nacha bí **craobh** 'na crann, do chí barr caomh os do chionn *Adv.* 72.2.2 16*r*. **1449-93** Freim na feile trén g^c^ tíre nior ér aoinfher no dáimh doiligh **craobh** fhial oinigh ó fhiadh noilagh nior fhás uime acht rioghna is ríogha *Bk. Clanranald* in *Rel. Celt.* II 264.

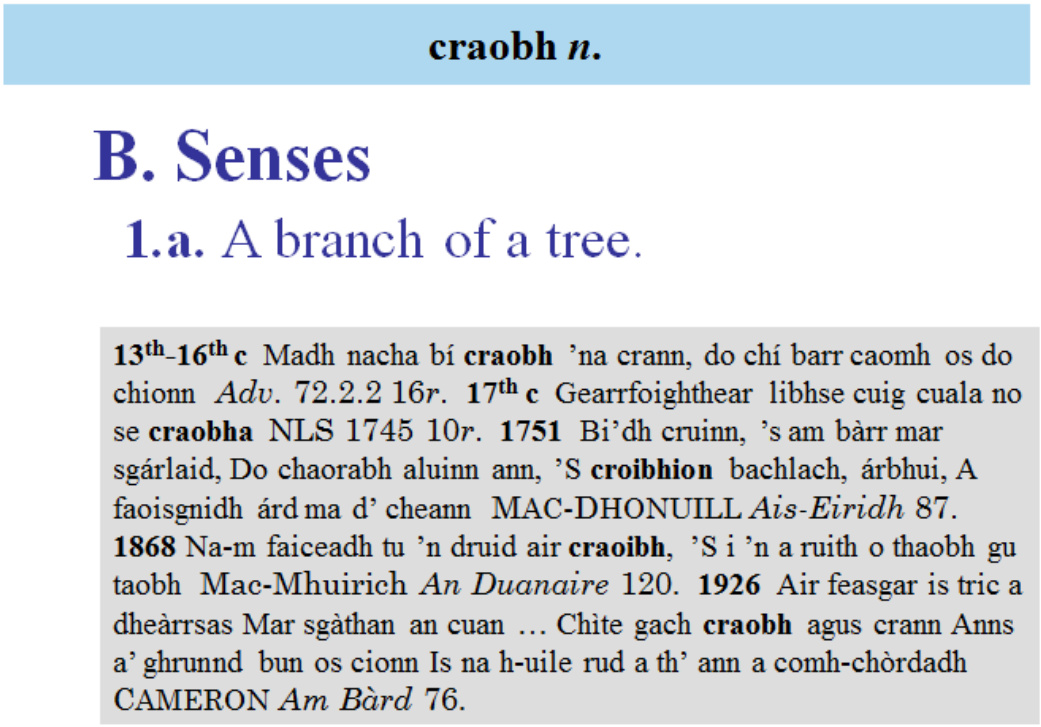
***Accusative*: 1567** Na Phairisidh, neoch ré bfaicthear an dadamh a suil a chumpanaigh agas nach bfaiceand **craobh** mhor ina tsuilibh féin Carswell 784. **1659** Mar **chraoibh** is amhluidh beithann se gcois aibhne fas ata, Do bheir na haimsir toradh trom, … Is soirbhidh leis gach ni da ndean *Psalms* 1:3.

***Vocative*: 1776** Ursainn-chatha Innse-gall, … Iuchair flatha na'm fior rann, **A Chraobh** sin a theasd do shiol Chuinn *Comh-chruinneachidh Orranaigh Gaidhealach* 24.

Illustration 2 – Section A Forms

Highlighting the orthographic focus separately in this way will leave Section B, the 'meat' of the dictionary, free to focus on the development of the senses with

quotations again providing first-hand evidence. Thus, material will be included in this section for semantic reasons only and will tell the story of the language and its people through the ages, beginning with the earliest evidenced senses and illustrating the semantic development of the word up to the present day.



craobh *n.*

B. Senses

1.a. A branch of a tree.

13th-16th c Madh nacha bí **craobh** 'na crann, do chí barr caomh os do chionn *Adv.* 72.2.2 16*r*. **17th c** Gearrfoighthear libhse cuig cuala no se **craobha** NLS 1745 10*r*. **1751** Bi'dh cruinn, 's am bàrr mar sgárlaid, Do chaorabh aluinn ann, 'S **croibhion** bachlach, árbhui, A faoisgnidh árd ma d' cheann MAC-DHONUILL *Ais-Eiridh* 87. **1868** Na-m faiceadh tu 'n druid air **craoibh**, 'S i 'n a ruith o thaobh gu taobh Mac-Mhuirich *An Duanaire* 120. **1926** Air feasgar is tric a dheàrrsas Mar sgàthan an cuan … Chite gach **craobh** agus crann Anns a' ghrunnd bun os cionn Is na h-uile rud a th' ann a comh-chòrdadh CAMERON *Am Bàrd* 76.

Illustration 3 – Section B Senses

Thus, the basic principle of dividing the language into 'building blocks' will enable users to access the information on the levels they want quickly and easily. It is hoped to include such additional options as: a list of synonyms; a phonological component giving guidance on pronunciation; illustrations where appropriate; and also links to other dictionaries and corpora.

Computerisation in historical lexicography allows sharpening of the focus in presentation in a way denied by the spatial limitations of the paper dictionary. It also opens up exciting possibilities in terms of the dictionary

resources that can be made available to the public almost simultaneously. One of the primary aims of the Faclair na Gàidhlig project is to maximise accessibility to the language. For this reason, the Gaelic-English historical dictionary will be compiled first. This could be viewed as the 'Gaelic *DOST*' in that it will define Gaelic using English as its target language. Like the *DOST*, or its latter stages at least, it will be defined from the perspective of its source language. This is extremely important since to express the meaning of words in a language accurately, the world must be viewed from its perspective. If this is done successfully, the monolingual and bilingual historical dictionaries will have the same structure. Only the meta-language will be different and users will be able to choose to access the dictionary through Gaelic or English. Those choosing to access *craobh* through Gaelic will see the following version of the dictionary:

craobh *ainm.*

Riochdan | Litreachadh | Freumhachadh | A Riochdan | B Tùsan

Às-earrannan | Co-fhaclan | Cairt bhliadhnachan | Fuaimneachadh | Dealbh

DASG | Tobar an Dualchais

craobh, *ainm.b. gin. sing.* **craoibhe**, *iol.* **craobhan**

craobh, *ainm.b.* ***Sing.*** *Ainmneach, cuspaireach*: **craoibh**. *Gairmeach*: **a chraobh**. *Ginideach*: **craoibhe, craobh, craoibh**. *Tabhartach*: **craoibh, craobh,** ***LDLM*** **creive.** ***Dùbailte*** **dá chraoibh.** ***Iol.*** *Ainmneach, cuspaireach* : **craobhan, craoibh, craobha, croibhion.** *Ginideach* : **craobhan, craobha, chraobhan, chraobh, craobh.** *Tabhartach* : **craobhan, craobha, craobhaibh, craobhuibh.**

[S.G. *cráeb*, *cróeb*, N.G.T. *craobh* geug, *corra uair*. craobh.]

Illustration 4 – Gaelic version of current forms, spellings and etymology

craobh *ainm.*

A. Riochdan

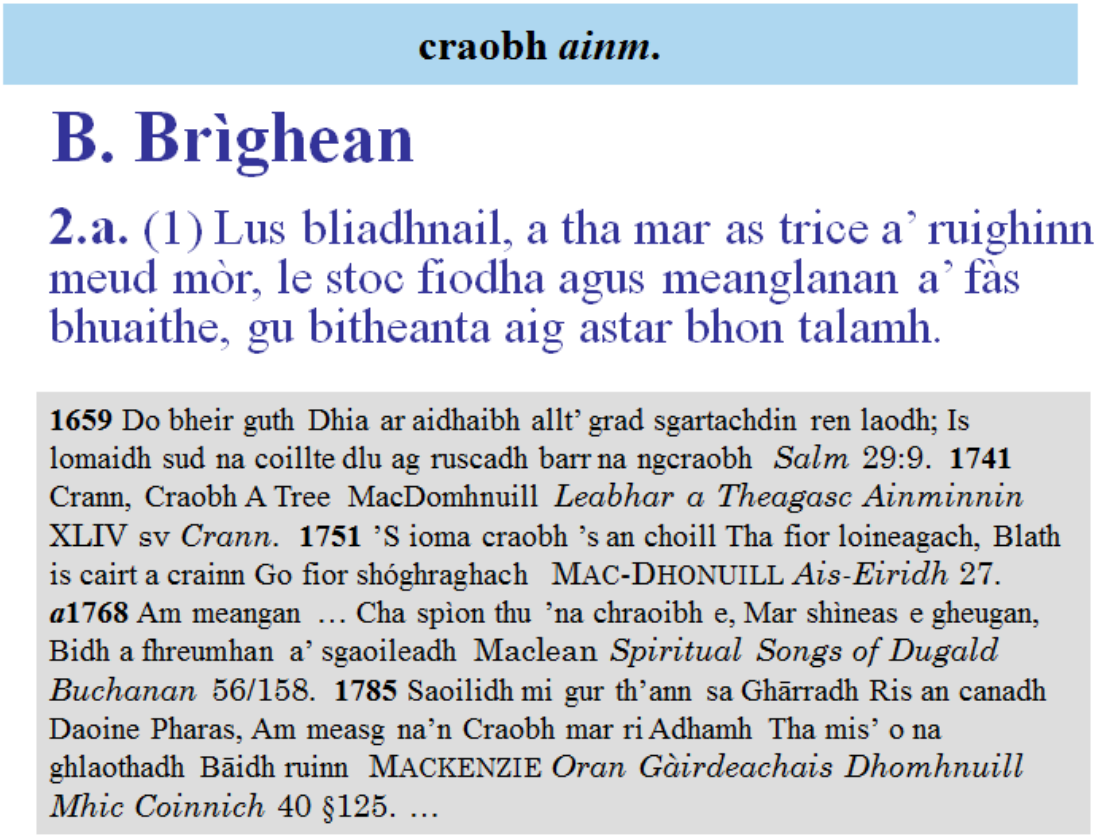
Singilte

***Ainmneach*: 13th-16th c** Madh nacha bí **craobh** 'na crann, do chí barr caomh os do chionn *Adv.* 72.2.2 16*r*. **1449-93** Freim na feile trén g^c^ tíre nior ér aoinfher no dáimh doiligh **craobh** fhial oinigh ó fhiadh noilagh nior fhás uime acht rioghna is ríogha *Bk. Clanranald* ann an *Rel. Celt.* II 264.

***Cuspaireach*: 1567** Na Phairisidh, neoch ré bfaicthear an dadamh a suil a chumpanaigh agas nach bfaiceand **craobh** mhor ina tsuilibh féin Carswell 784. **1659** Mar **chraoibh** is amhluidh beithann se gcois aibhne fas ata, Do bheir na haimsir toradh trom, … Is soirbhidh leis gach ni da ndean *Salm* 1:3.

***Gairmeach*: 1776** Ursainn-chatha Innse-gall, … Iuchair flatha na'm fior rann, **A Chraobh** sin a theasd do shiol Chuinn *Comh-chruinneachidh Orranaigh Gaidhealach* 24.

Illustration 5 – Gaelic version of Section A Forms



craobh *ainm*.

B. Brìghean

2.a. (1) Lus bliadhnail, a tha mar as trice a' ruighinn meud mòr, le stoc fiodha agus meanglanan a' fàs bhuaithe, gu bitheanta aig astar bhon talamh.

1659 Do bheir guth Dhia ar aidhaibh allt' grad sgartachdin ren laodh; Is lomaidh sud na coillte dlu ag ruscadh barr na ngcraobh *Salm* 29:9. **1741** Crann, Craobh A Tree MacDomhnuill *Leabhar a Theagasc Ainminnin* XLIV sv *Crann*. **1751** 'S ioma craobh 's an choill Tha fior loineagach, Blath is cairt a crainn Go fior shóghraghach MAC-DHONUILL *Ais-Eiridh* 27. ***a*1768** Am meangan … Cha spìon thu 'na chraoibh e, Mar shìneas e gheugan, Bidh a fhreumhan a' sgaoileadh Maclean *Spiritual Songs of Dugald Buchanan* 56/158. **1785** Saoilidh mi gur th'ann sa Ghārradh Ris an canadh Daoine Pharas, Am measg na'n Craobh mar ri Adhamh Tha mis' o na ghlaothadh Bāidh ruinn MACKENZIE *Oran Gàirdeachais Dhomhnuill Mhic Coinnich* 40 §125. …

Illustration 6 – Gaelic version of Section B Senses

Computerisation will also allow lexicographers to feed definitions into databases to enable production of derived resources in tandem, for example, single volume dictionaries for Gaelic-medium education, dictionaries for learners, and so on. The thought processes involved to understand a word fully need only be gone through once and a variety of resources can be produced with much less effort and expense than in the past.

2.1 Faclair na Gàidhlig Foundation Project

A project of this magnitude and importance obviously requires meticulous planning and preparation before dictionary compilation can begin. One of the most important challenges that must be met is how to create an historical lexicographical tradition quickly and effectively for a language where there is none. Traditionally, such projects began with the creation of a slip archive, usually by volunteers writing out quotations they themselves had chosen. This stage in itself took decades to complete. Material was then pre-edited and subsumed under headwords, subsequently analysed and dictionary entries compiled. Staff learned the skills of lexicography in-post, working closely with an experienced colleague. Publication was in fascicles in alphabetical order and editorial policy was developed and refined as the project progressed, not reaching its peak until at least halfway through the alphabet, thus ensuring that a significant proportion of the early part of the dictionary would require upgrading when the end of the alphabet was reached. It was obvious at the start of Faclair na Gàidhlig that this methodology was unsustainable in the context of the present day.

Clearly, a new approach was needed to enable Gaelic lexicographers to produce the dictionary as quickly as possible without compromising quality. A detailed planning stage would be necessary to ensure that the project would progress smoothly and expeditiously and attract continued funding in order to sustain steady production. That planning stage is now well underway and has four essential components. The first of these is the editorial foundation. In this part of the project, sample entries are produced for the dual purpose of determining the lexicographical structures most suited to the language, and of producing detailed instructions for compiling these sample entries in order to train the first historical lexicographers in Gaelic. These instructions currently range from 24 to 212 pages for each word and provide templates for the 94 sample entries so far compiled. Every thought process, all the assessment and reassessment in analysis, every decision made and all the detail of method of presentation is written in these instructions, from noting the earliest example of a spelling form by marking the slip with an F (indicating that the form must be included in Section A), to writing the actual definition which will appear in the dictionary. These instructions will guide trainees, step by step, from quotation

slips to final entry simulating the conventional method of learning in-post alongside an experienced lexicographer. Indeed, a significant advantage of this new approach is that they could accelerate progress of the project by making it possible to train several lexicographers simultaneously.

The second essential component of planning is the textual foundation, the body of evidence selected from the language on which dictionary will be based. It is crucial that this is representative of all the evidence available to us, particularly in terms of date, register and dialect area. When the Faclair na Gàidhlig project was established in 2003, the National Library of Scotland Gaelic catalogue contained around 3,000 published titles. Professor Donald Meek, who was one of the great driving forces behind the establishment of the project, whittled this down to a list of 205 key printed texts which will form the basis of the evidence from which the dictionary will be created. These 205 texts, covering the language from the late sixteenth century to the present day, were assessed in terms of their usefulness to the dictionary by Dr Catriona Mackie who was the Leverhulme-funded Research Assistant from 2005–08. She compiled reports for each text, averaging 3-4 pages each in length, and giving information on such aspects as register, style, geographical origin, social context, language date and bibliographical details, providing new lexicographers with instant access to information on texts at a level usually only acquired after some years in post. Manuscript material, covering the period from the twelfth century to the nineteenth, will also be included and their digitisation will form a separate part of the project, although fortunately an increasing number are becoming available in digital format.

The textual foundation feeds into the third essential component of planning, the digital foundation which will be explained in greater detail in §3 below. This will give lexicographers 'fingertip access' to the textual foundation and the digitised texts will be supported by customised versions of the written reports. The development of software for dictionary compilation is also part of the digital foundation and will include a system of electronic excerpting, ways of manipulating material during the compilation of entries and a system for editors' notes to each other. Paper slips will continue to be generated for larger entries as long as this remains the most satisfactory way to manipulate a large amount of material. After the entries are compiled additional searches can be

done on the much more extensive DASG corpus to fill in any perceived gaps in coverage. The final component of the digital foundation will be the user-interface for the dictionary itself.

The fourth essential component of project planning is human resources both in terms of the operational management of the project and the compilation of the dictionary. The current project structure comprises a management committee, a project team and an advisory board. Faclair na Gàidhlig has been very fortunate to have had since its inception a strong management group in the form of a Steering Committee composed of representatives of the partner institutions who, as Gaelic academics, are fully committed to its creation. The development of the editorial foundation for the dictionary is the remit of the project team. The current team consists of one full-time member of staff, Lorna Pike, who was formerly one of the editors of the *DOST* and three consultants: one Lexicographical Consultant – Marace Dareau, former Editorial Director of the *DOST*; and two Language Consultants – Professor Emeritus William Gillies, Honorary Professorial Research Fellow at the University of Edinburgh, and Emeritus Professor Colm Ó Baoill, Emeritus Professor of Celtic at Aberdeen University. This interdisciplinary team aims to co-ordinate its knowledge and import skills to Gaelic that are essential for the creation of a high quality historical lexicographical resource. Quality control is exercised by members of the Advisory Board for the project, drawn from the fields of lexicography and Celtic Studies in Scotland, Ireland and Wales, and who review outputs annually. The lexicographers who will undertake to compile the dictionary will be honours graduates in Celtic and Gaelic Studies. This is a relatively small pool from which to recruit personnel for a nationally important project and careful planning is necessary to ensure that suitably qualified people will be available to train in the profession. Celtic and Gaelic departments in Scottish universities have been urged to consider the requisite skills in course-planning, and courses have been developed in a number of pertinent areas of linguistics and sociolinguistics.

2.2 Creating the lexicographical tradition

An integral part of creating the lexicographical tradition is the editorial foundation; the fact that work could begin work on this aspect of Faclair na Gàidhlig before the digital corpus was complete has been a significant factor in its establishment and sustainability. The raw materials, i.e. the quotations used in the 94 sample entries, are those produced by the Historical Dictionary of Scottish Gaelic project begun by the late Professor Derick Thomson at Glasgow University. (Macdonald [1987] 1994: 62–63; Ó Maolalaigh 2008c: 474–76). This project was underfunded and understaffed and was suspended in the late 1990s with no published lexicographical output. However, an archive of around 550,000 slips, amongst other resources including fieldwork collections and sound recordings, was created and it has proved crucial to the progress of Faclair na Gàidhlig. Indeed had it not been available, it would not have been possible to undertake any work on the editorial foundation until a corpus had been produced. By that time, the skills of working from first principles in historical lexicography would be gone from Scotland and it is difficult to envisage how such a project could have ever been established for Gaelic.

The Historical Dictionary of Scottish Gaelic Archive (HDSG-A), now forming part of DASG, is probably a third of the size needed to produce an historical dictionary. Furthermore, the paper slips themselves contain quotations which are, on the whole, too brief to enable any sound lexicographical analysis. However, the archive, with some augmentation, has been more than adequate for investigating lexicographical structures and promises to be sufficient for use in teaching the skills of historical lexicography. Its slips fall into four categories: handwritten slips from printed texts; handwritten slips from manuscripts; computer-generated slips from the Old Testament and verse, and handwritten slips transcribed from fieldwork with oral informants in Scotland and Canada between 1966 and *c*. 1992. The last category was omitted from the sample entries for data protection reasons. However, these slips almost certainly contain material that is now lost to the spoken language and every effort will be made to make use of them in Faclair na Gàidhlig. The computer-generated slips are amongst the most useful since, in most instances, they provide enough material to work with. On the other hand, the handwritten slips,

particularly from printed texts, tend to be too brief and require a great deal of augmentation and checking to make them useful. The following example demonstrates that a considerable amount of material has had to be added to the original quotation which is underlined.

craobh-fheòir Ch√
0001214

O chionn ghoirid thill Mr. D. MacAula, ball agus fear dreuchd sa chomunn gu
A short time ago Mr D. MacAula, a member and office bearer in the society,
dhachaidh dhuthchasaich sa Ghaidhealtachd, agus beagan mun d' fhalhh [*sic*] e,
returned to his native home in the Gaidhealtachd, and a little before he went away,
thugadh dha bata craobh fheoir (*grass tree*) *New Zealand* air a dheanamh
he was given a boat of New Zealand grass tree (grass tree) decorated with
sgiamhach le or, le deadh ruin a chomuinn;
gold, with the sincere affection of the society;

Source: Mac-Talla, vol VIII, no. 13 Date 6/10/99 [Some damage to page. – LP]
Notes: [Uilleam Mac Leoid] ("g. tree" e "N. Zealand" italicised in text)

1899 *Mac-Talla* VIII No. 13 100/4.

The original quotation 'thugadh dha bata craobh fheoir […] New Zealand' ('he was given a boat of New Zealand grass tree') is not very informative. There is no indication as to who 'he' was, why he was given the boat or, anything about the boat, other than the fact it was made of grass tree. The augmented slip shows that this was a leaving gift, decorated with gold and was most likely a model boat. This information would be key to defining *bàta* ('a boat'). This slip was in fact the only evidence for the compound word *craobh-fheòir* ('grass-tree'). Often words with scanty evidence take a great deal of work in comparison to those with many examples and this one was no exception. The instructions for analysing the sense of this quotation take up about three quarters of a page and provide a brief insight into the working life of a lexicographer. The obvious thing to notice first is that this is not a reference to

the tree itself but to its wood as a commodity to be made into something. Further research in order to try to identify the type of tree involved revealed that the grass tree grows only in Australia and in this case the context is New Zealand. In the *OED* which is generally reliable for tree species, the definition of 'grass tree' also includes the lancewood and the cabbage tree of New Zealand. Further investigation showed that the lancewood has tough timber and that the cabbage tree has a fibrous trunk. The lancewood was therefore the more likely candidate to be used for making a boat. This is reflected in the definition: 'The wood of the grass tree. Since the context is New Zealand, the tree referred to is probably the lancewood, *pseudopanax crassifolium*.' So this sample entry, with only one example, has proved quite educational to the trainee lexicographer in terms of the research required to write the definition. To date, over 1,800 of such slips have been used in sample entries, each one passing through five stages of semantic analysis: (i) reading the slips and making initial sense divisions, (ii) developing the sense structure, (iii) adjusting the senses and writing definitions, (iv) selecting the quotations to be included and the material to be quoted, and, finally, (v) refining the definitions. Thus, over the 94 sample entries, these five processes give at least 9,000 bytes of information to deal with and often there is more than one fact to be learned from a slip at any one stage. Thus, the HDSG-A has been of immense value in the foundation stage of Faclair na Gàidhlig and a great debt of gratitude is owed to all those, most of them volunteers, who were involved in its creation. But as far as a dictionary beginning in the twenty-first century is concerned the only way forward is to begin again with a new digital corpus of the language.

3 Corpas na Gàidhlig

As noted above (§1), one of the initial aims of Corpas na Gàidhlig is to provide the digital textual corpus upon which Faclair na Gàidhlig will be based. It also has the longer term aim of providing the first comprehensive freely available online textual corpus for the Gaelic language which will inform future research and provide the basis for the development of technological, pedagogic and reference resources for the language. As such it aims to be at the heart of future corpus planning projects for Gaelic. Once the textual corpus for Faclair na Gàidhlig has been established, the intention is to expand Corpas na Gàidhlig to

embrace as wide a range as possible of different registers, genres, modes and styles to eventually include speech, drama, private letters, formal minutes of meetings and so on. In terms of speech, we are very fortunate in having as part of the DASG archive valuable tape recordings from throughout Scotland and parts of Nova Scotia. These were recorded as part of the former Historical Dictionary of Scottish Gaelic project. These recordings were fully digitised during 2011 and it is intended that these will eventually be added to Corpas na Gàidhlig.

3.1 Textual corpus for Faclair na Gàidhlig[1]

A total of 205 printed texts have been identified as the initial core textual basis for Faclair na Gàidhlig. The vast majority of the printed texts date from the mid eighteenth to the end of the twentieth century. The earliest printed text is the first book ever to be published in Gaelic, *Foirm na n-Urrnuidheadh*, John Carswell's Gaelic translation of the Book of Common Order which was published in 1567 (Thomson 1970). A range of manuscript texts will also be included, the earliest of which is the twelfth-century Gaelic Notes contained in the Book of Deer (Jackson 1972; Ó Maolalaigh 2008a; Forsyth, Broun and Clancy 2008). The most recent text to appear in the Corpus thus far is Iain MacLeòid's 2005 novel *Na Klondykers*.

3.1.1 Register, genre and mode

A working classification of the genres of the 205 texts recognises nine categories and is presented in table 1.

[1] This section draws on an article by Ó Maolalaigh (2013), which contains more detailed information on the textual corpus being compiled for Faclair na Gàidhlig.

Genre	**Description**
Public letter	letters published in a newspaper or periodical
Proverb	proverbs, riddles, idioms
Administrative	legal, business, parliamentary, regulatory, language planning
Biography	biography and autobiography
Miscellaneous	a range of genres
Imaginative	novels, short stories, narratives, modern (non-traditional) poetry
Expository	description, analysis, classification, instruction
Religious	sermons, catechisms, translations
Traditional	oral narrative, traditional song and verse

Table 1 – Genres of texts

The ascription of category is not always straightforward as there is naturally a certain amount of cross-over between certain categories, which is captured by the category of 'miscellaneous'. Statistics relating to genre and mode (i.e. prose and verse) are given in Table 2 and Figure 1. While prose is the predominant mode, a high percentage of verse texts is included in the corpus. This reflects very much the literary tradition of Gaelic as reflected in the published record.

Genre	Prose	Verse	Prose and Verse	Total	%
Public letter	1	0	0	1	0.5%
Proverb	3	0	1	4	2.0%
Administrative	6	0	0	6	2.9%
Biography	5	0	1	6	2.9%
Miscellaneous	1	0	14	15	7.3%
Imaginative	21	4	3	28	13.7%
Expository	25	1	3	29	14.1%
Religious	28	16	6	50	24.4%
Traditional	3	58	5	66	32.2%
Total	**93 (45%)**	**79 (39%)**	**33 (16%)**	**205**	

Table 2 – Genre and mode

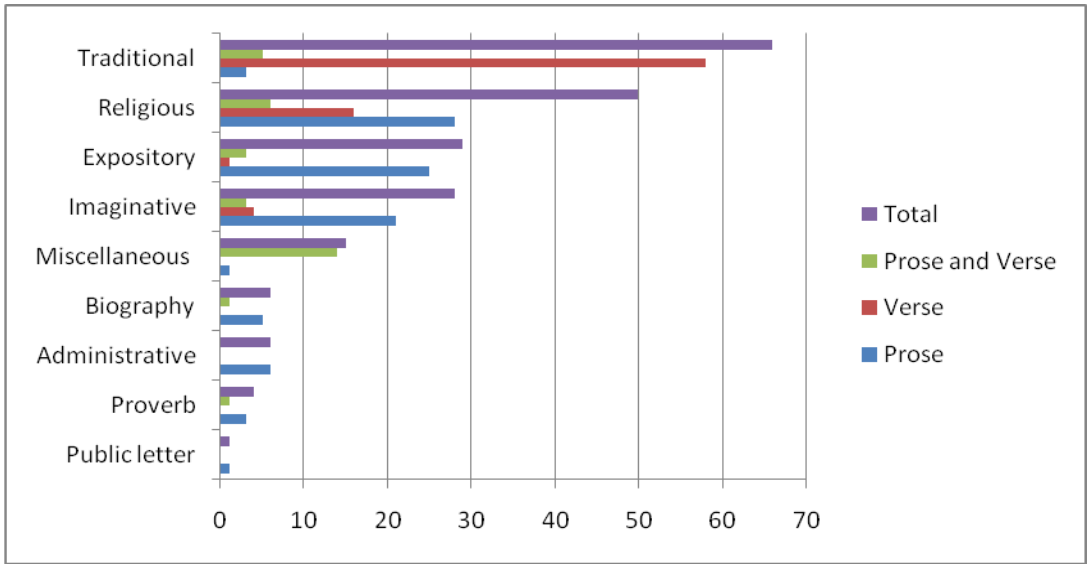


Figure 1 – Genre and mode

It can be readily seen that the genres most represented are traditional (32.2%) and religious (24.4%). These two categories combined contain the vast majority of all verse texts in the corpus, i.e. 94%. Traditional texts are overwhelmingly represented by verse, thus illustrating that traditional prose oral narrative is under-represented in the corpus as currently envisaged. Although five major periodicals from the nineteenth century appear, none of the twentieth-century magazines, papers or periodicals are currently included, such as *Gairm*, which played a major role in twentieth-century Gaelic literature and publication (MacilleDhuibh 2008). A selection of these and other textual types will be added to the corpus at a later stage.

3.1.2 Geographical origin

In terms of geographical location, the majority of texts represent the Gaelic of the outer and inner islands, although there is a fair representation of some mainland areas as seen in Figure 2.

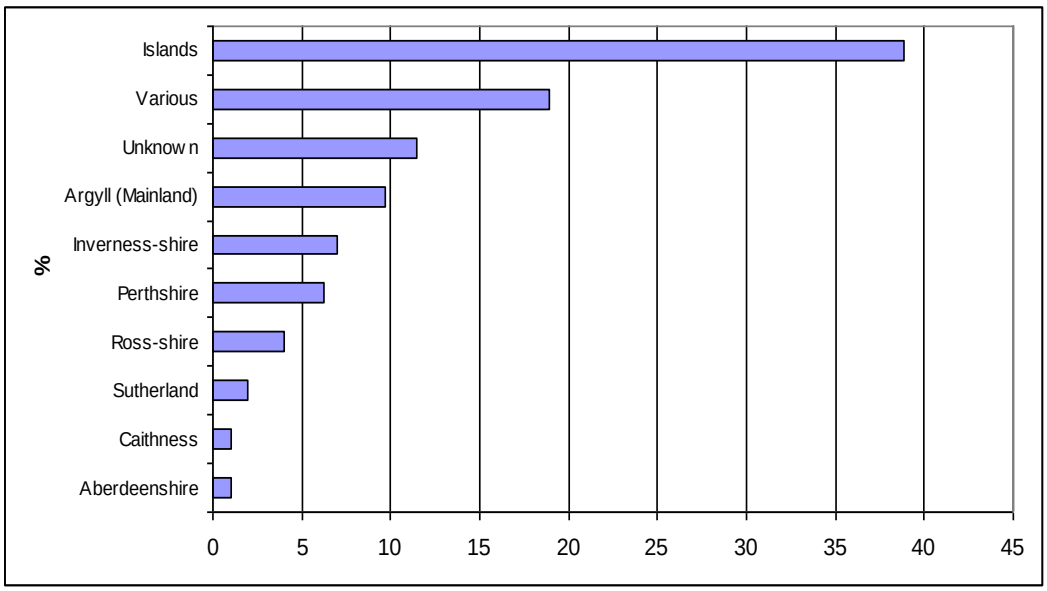


Figure 2 – Geographical origin

Figure 3 illustrates that the Isle of Lewis is the island most represented in the corpus.[2]

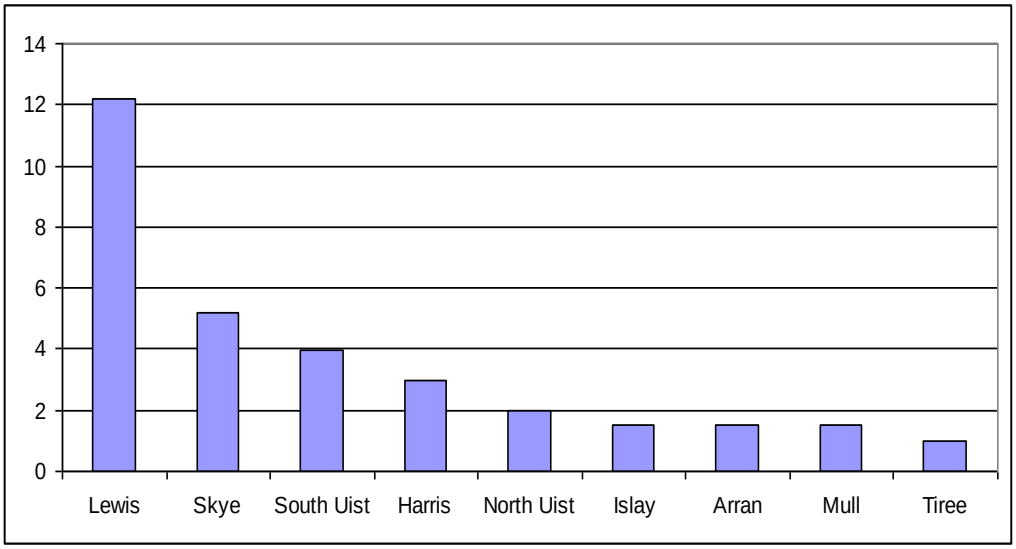


Figure 3 –Islands represented in the textual corpus

3.2 Current progress

Work to date has concentrated on capturing in digital format the 205 texts identified for the first phase. This work involves 5 discrete stages:

1. The first step is to identify whether a particular text has been captured digitally by other projects or individuals. The Internet Archive (archive.org) resource has been particularly helpful in this regard. Where digital texts exist these are accessed – with the appropriate permissions from the host or donor.

2. Where digital images do not exist, texts are scanned using a standard scanner.

[2] The Advisory Board is addressing the imbalances identified in graphs 1, 2 and 3.

3. Scanned texts are then OCRed using OmniPage Professional 17 and saved as ASCII files.

4. All texts are then edited to reflect as far as possible the original format of printed pages.

5. Texts are proofread twice by different individuals.

All activity is logged carefully and dated for quality assurance purposes.

The project has employed a number of full-time, part-time and casual corpus workers over the past few years. Two corpus assistants (1.5 FTE) are currently (July 2012) employed: Olga Szczesnowicz (Senior Corpus Assistant) from Poland and Linda MacLeod (Corpus Assistant) from North Uist – both graduates in Celtic and Gaelic. Stephen Barrett has recently been employed as IT Systems Developer for DASG and Faclair na Gàidhlig.

Progress to date (July 2013) is summarised in Table 3. So far a total of 145 texts have been through the digitisation process, from scanning to being proofread twice. This provides nearly seven million words of what may be referred to as ‘corpus-ready’ text.

	Scanned or in Digital Form	**Edited**	**OCRed**	**Proofed**	**Proofed twice**	**No. of Words**
Corpus-ready texts (proofed twice)	145	145	145	145	145	6,856,011
Proofed texts	1	1	1	1	–	160,634
Digitised but requiring proofreading	0	0	0	–	–	–
Digitised but requiring OCR & proofreading	0	0	–	–	–	–
Texts available on web	33?	–	–	–	–	–
TOTAL	**179**	**146**		**146**	**145**	**7,016,645**
TOTAL: % of total 205 texts	**87%**	**71%**		**71%**	**71%**	

Table 3 – Progress to date (1 July 2013)

3.3 Digital foundation

It is too early to report on technical aspects of the corpus search engine, which is currently under development utilising Open Corpus Workbench (CWB), a collection of open-source tools for managing and querying large text corpora. IT work has concentrated on developing a web-based lexicographical system, with the first stage of development concentrating on the production of virtual slips containing quotations for individual words, thus replicating the traditional system of producing 'paper' slips, which can be manipulated and edited

onscreen in a variety of ways. The features of the virtual slips are illustrated in what follows by a series of seven screenshots and accompanying commentaries.[3]

3.3.1 The search page (Screenshot 1)

The interface for the system is bilingual, enabling future lexicographers to access and interrogate the data in Gaelic or English. The English interface is used in this paper for the purposes of illustration.

Screenshot 1 – The search page

[3] The authors are grateful to Stephen Barrett for providing these screenshots, all of which were created in July 2012. We would also like to acknowledge here the valuable input of Dr Mark McConville, Soillse Research Fellow at the University of Glasgow, who has been co-supervising the technical IT developments of the project.

3.3.2 Word searching (Screenshot 2)

A word search (conducted in July 2012) for the word *craobh* ('tree', etc.) provides 1,170 results.

Screenshot 2 – Word search for *craobh*

Each line represents an abbreviated form of a unique slip and contains five columns or fields, consisting of: (i) the short title of the source containing the word, (ii) the author or editor of the source, (iii) the page number of the source where the word is located, (iv) the full slip (which is explained further below) and (v) a short extract containing the (highlighted) search word in context.

The default search includes lenited (i.e. forms with *ch*-) and palatalised (slenderised, i.e. inflected forms with –*ibh*(-)) forms, which are highly prevalent in Gaelic, i.e. *chraobh*, *craoibh*, *chraoibh*, *craoibhe*, *chraoibhe*.[4] This maximises the number of hits relevant to the lexicographer. These features as well as the the accent feature can be easily turned on and off when refining searches.

3.3.3 Summary of short title details (Screenshot 3)

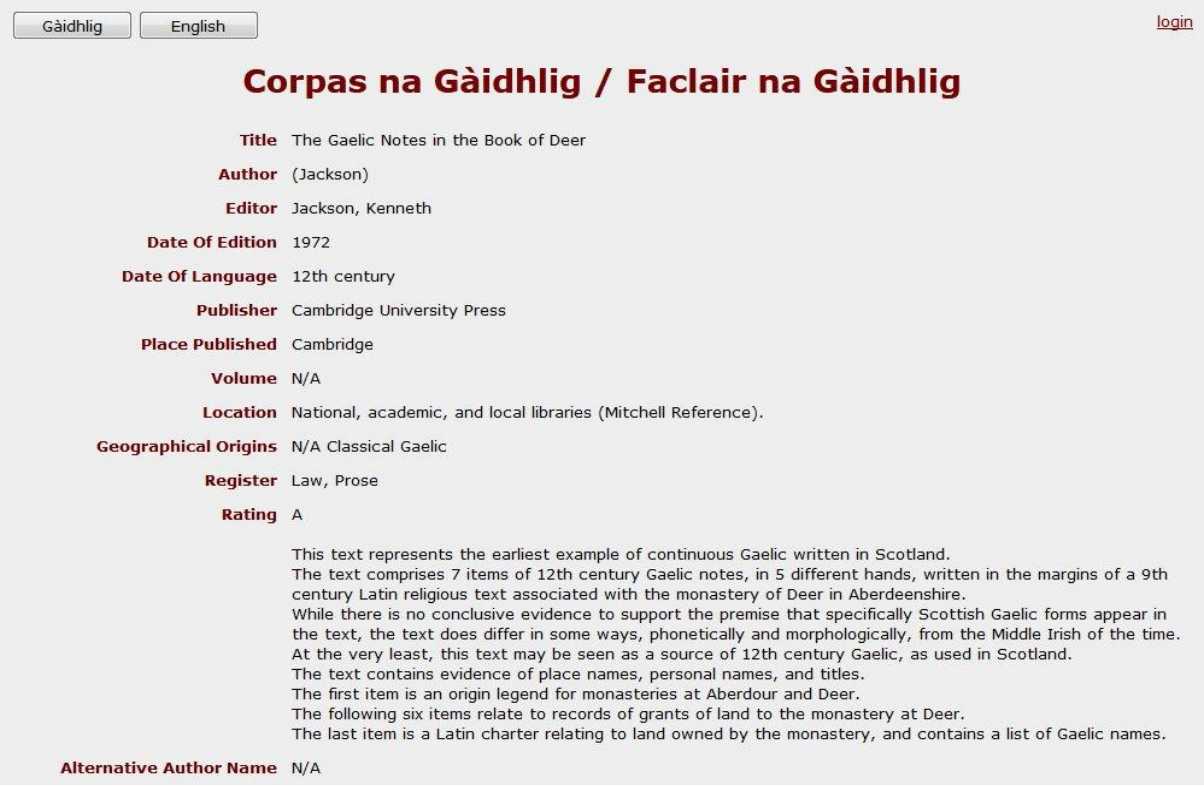
Hovering the mouse over the short title in any of the results produces a rectangular box which displays brief summary details of the metadata associated with the source in question. At present this includes 3 fields of information, namely (i) date of language, (ii) register and (iii) geographical origin.

[4] For a brief introduction to the initial mutation 'lenition' and the morpho-phonological process of palatalisation (traditionally referred to as slenderisation), see Ó Maolalaigh (2008b: xxii–xxiii).

Screenshot 3 – Summary of short title details

3.3.4 Full metadata relating to source (Screenshot 4)

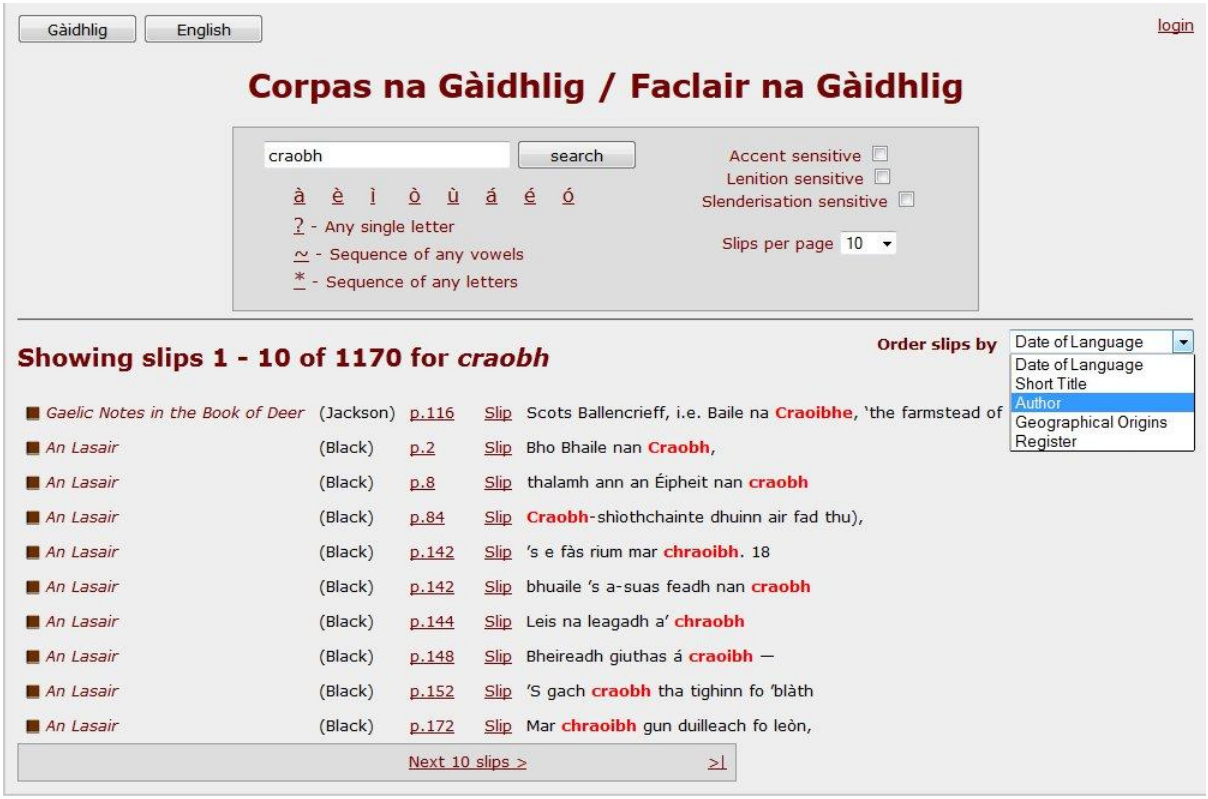
Clicking the box to the left of the short title retrieves all of the metadata associated with the source. This includes: (i) all bibliographic information, (ii) date of language, (iii) register and other details such as a discursive commentary on the contents of the source in question. These details represent customised versions of the reports prepared by Catriona Mackie as part of the earlier Leverhulme project (2005–08) referred to in §2.1. It is planned to regularise the structure of this data to maximise the searchability of their content.



Screenshot 4 – Full metadata relating to source

3.3.5 Arrangement of search results (Screenshot 5)

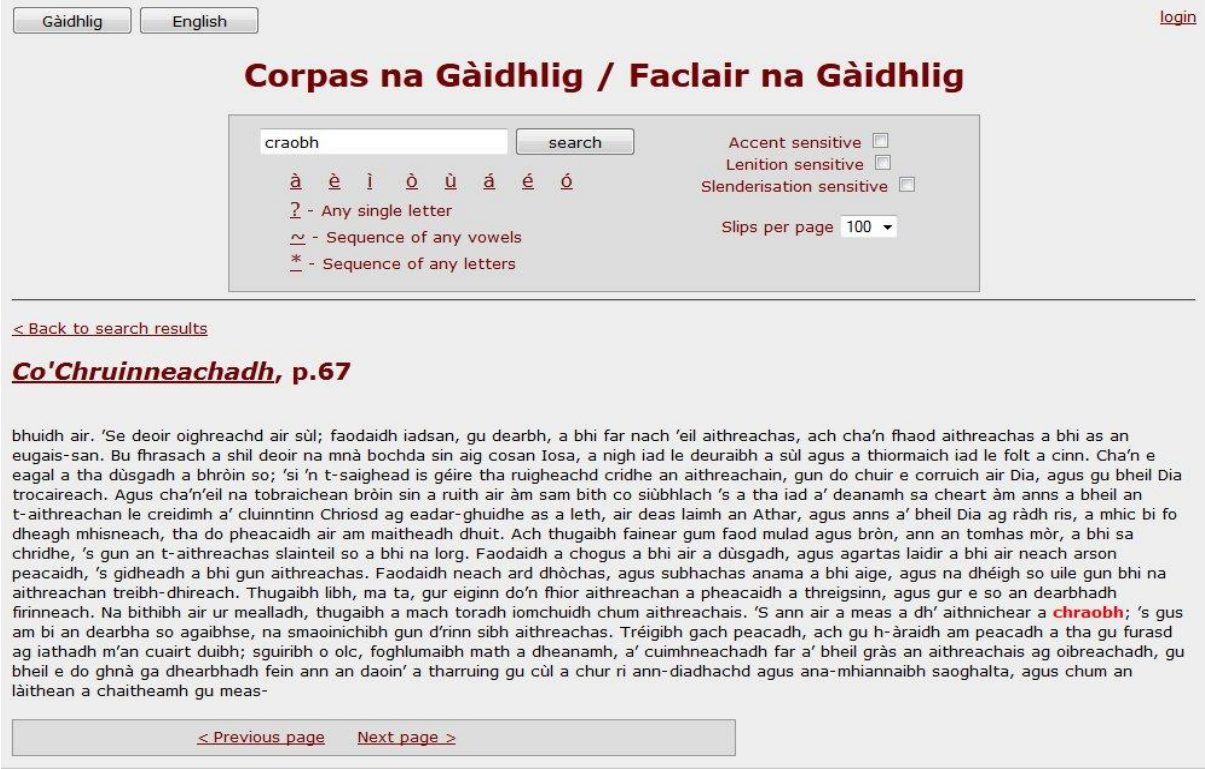
The results of any search can currently be arranged according to: (i) date of language, (ii) short title, (iii) author, (iv) geographical origins and (v) register.



Screenshot 5 – Arrangement of search results

3.3.6 The page context (Screenshot 6)

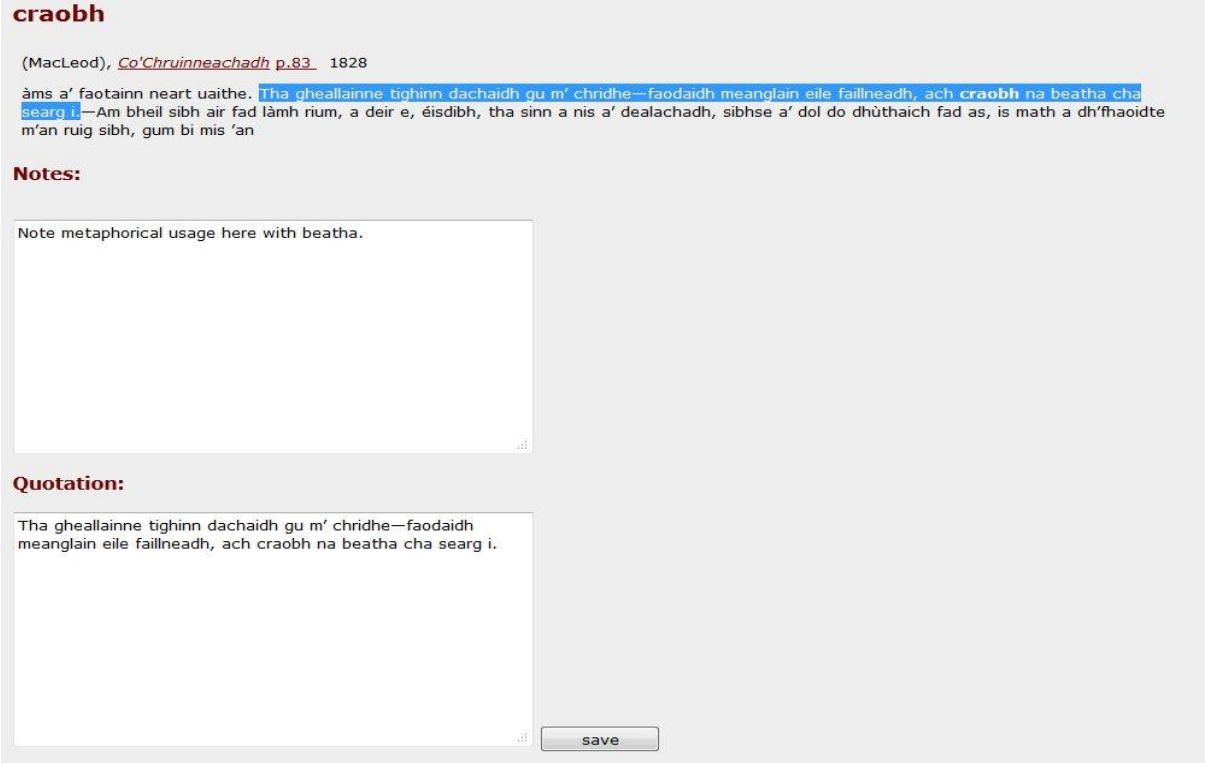
Clicking on the page field brings the user to the full page of the source in which the search word occurs, thus enabling lexicographers to view the full context of the search word, which is highlighted.



Screenshot 6 – The page context

3.3.7 The slip (Screenshot 7)

Clicking the slip field produces the full virtual slip which contains the main details about the source, i.e. author, short title, page number and date of publication. In addition to this the search word is cited in context. There are also two editable fields for lexicographers: (i) a notes field for sundry notes and (ii) a quotation field. The quotation is the quotation that will ultimately be used in the final dictionary. This can be cut and pasted from the citation and edited if necessary. Changes to the slip can be saved, with the editor's name and date of change recorded automatically.



Screenshot 7 – The slip

Further developments and enhancements of the lexicographical system are envisaged, including the development of an interface which will enable lexicographers to produce the dictionary entries based on the virtual slips.

4 Case study: Gaelic numerals '3' to '10' + singular nouns

4.1 Corpus

For this brief case study, a relatively small corpus of 1.5 million words has been assembled, representing a selection of 57 texts from the Corpus.[5] The modes and genres represented in this sub-corpus are described in Table 4.

Mode	**No. of texts**	**%**
Prose	39	68.4%
Verse	8	14.0%
Prose & Verse	10	17.5%
Genre		
Imaginative	21	36.8%
Expository	11	19.3%
Traditional	11	19.3%
Proverbs	1	17.5%
Religious	1	17.5%
Biography	5	8.8%
Miscellaneous	4	7.0%
Administrative	3	5.3%
Total	57	100%

Table 4 – Mode and genre of sub-corpus

The software used in order to observe patterns in the corpus is the freely available AntConc concordance package, version 3.2.1 for Windows, developed by Laurence Anthony of Waseda University in Japan (Anthony 2005).[6]

[5] This case study draws on Ó Maolalaigh (2013), where a more detailed discussion appears.

4.2 Background

General accounts of Scottish Gaelic grammar inform us implicitly or explicitly that the cardinal numerals from 'three' to 'ten' are followed by the plural form of the noun they qualify, e.g. *trì coin* ('three dogs'), *ceithir cait* ('four cats'), *deich leabhraichean* ('ten books') but the complexities surrounding these numerals tend not be confronted (see, for instance, Ó Maolalaigh 2008b: 112–13). It is rarely mentioned that the singular form of nouns can be used with the numerals from 'three' to 'ten' or that lenition can follow the numerals 'three', 'four' and 'five'. Some accounts provide minor additional details but we still lack a full and comprehensive account of the grammar of numerals in Scottish Gaelic.

Professor David Greene, in his magisterial 1992 description of numerals in all of the Celtic languages, says of Scottish Gaelic that 'the only nouns which show singular forms after these numerals [i.e. 'three' to 'ten'] are the substantival numerals *fichead* "twenty", *ceud* "hundred", and *mìle* "thousand", together with the enumerators *dusan* "dozen", *duine* "person", *latha* "day" and *bliadhna* "year".' This provides a total of seven nouns which are regularly used in the singular with the numerals 'three' to 'ten'. However, an analysis of the corpus provides a much fuller picture of this aspect of Scottish Gaelic grammar. In particular, the corpus reveals that the singular is found with a total of forty nouns, two of which are English words ('fathom' and 'kilometre'). A comparison of Greene's list of seven nouns with the forty nouns appearing in the corpus illustrates very clearly the potential power of textual corpora to transform existing traditional descriptions of Gaelic grammar and to utterly revolutionise our understanding of Gaelic language.

4.3 Data from the sub-corpus

The fourteen most commonly occurring nouns in the singular with the numerals 'three' to 'ten' occurring in the corpus are listed in Table 5, which includes only nouns occurring more than six times in the singular with these numerals. Some of these nouns also occur in the plural with numerals, and this is illustrated in the third column of the table. These include all of Greene's examples except

[6] This software can be downloaded freely from *Laurence Anthony's Website* <http://www.antlab.sci.waseda.ac.jp/software.html> (accessed December 2012).

dusan which occurs only once with a numeral in our sub-corpus (*ceithir dusan* 'four dozen').

Noun	**Frequency of singular form with numbers 'three' to 'ten'**	**Plural forms attested with numerals 'three' to 'ten'**
bliadhna ('year')	436	*bliadhnachan* (5), *bliadhnaichean* (2), *bliadhnan* (1)
fichead ('twenty')	190	*ficheadan* (1)
mìle ('thousand', 'mile')	166	–
ceud ('hundred')	140	–
latha / *là* ('day')	115	*làithean* (16)
duine ('man, person')	25	*daoine* (1)
sgillinn / *sgilling* ('penny')	26	–
nota ('[pound] note')	12	*notaichean* (27), *notachan* (3), *notan* (1)
tastan / *tasdan* ('shilling')	12	*tastain* / *tasdain* (13)
cairteal / *cairteil* ('quarter')	8	–
tunna / *tonna* ('ton')	8	–
slat ('yard [measurement]')	7	*slatan* (7), *slata* (1)
oidhche ('night')	7	*oidhcheannan* (3), *oidhchean* (1)
muillean ('million')	6	–

Table 5 – The fourteen most commonly occurring nouns in the singular with the numerals 'three' to 'ten'

It is no coincidence that the five nouns which occur most commonly in the singular with the numerals 'three' to 'ten' (*bliadhna*, *fichead*, *mìle*, *ceud*, *latha* / *là*) are among the top six most commonly occurring collocates of the numerals. Their high frequency in the language helps to retain their irregular status when compared with the vast majority of other nouns.

If we combine the corpus data with data from other grammatical and dialect sources, we can identify a total of fifty nouns which are used in the singular with the numerals 'three' to 'ten'. The fact that 40 of these are present in our sub-corpus validates the usefulness and power of even relatively small textual corpora.

4.3.1 Classification

The vast majority of these fifty nouns represent a well-defined semantic / grammatical class of enumerators and quantifiers. These can be further categorised into seven sub-classes of: (i) numeral, (ii) time, (iii) length / distance, (iv) measure, (v) weight, (vi) currency, (vii) people. In addition to these we can recognise two further classes, namely (viii) a class of collective nouns and (ix) another class of abstract nouns – a category which arguably also describes all numbers, enumerators and quantifiers.

Each of the fifty words (with only one exception) can be assigned to one of these nine classes. Note that some words can be classified under more than one heading, e.g. *cairteal* (meaning 'quarter') can be a numeral or can relate to time.

(i) Numeral

ceàrn ('quarter'), *cairteal* / *cairteil* ('quarter'), *ceathramh* ('stanza', lit. 'quarter'), *ceud* ('hundred'), *ceudamh* ('hundredth'), *còigeamh* / *còigeadh* ('province', lit. 'fifth'), *dusan* ('dozen'), *fichead* ('twenty'), *ficheadamh* ('twentieth'), *mìle* ('thousand'), *mìleamh* ('thousandth'), *muillean* ('million'), *naodhnar* ('nine people'), *ochdamh* ('eighth')

(ii) Time

aois ('age [of animal]'), *bliadhna* ('year'), *cairteal* / *cairteil* ('quarter'), *cuairt* ('time', lit. 'round'), *latha* / *là* ('day'), *mionaid* ('minute'),

oidhche (‘night’), *seachdain* (‘week’), *tràth* (‘time, day’), *uair* (‘hour, time’)

(iii) Length / distance

aitheamh / *faitheamh* (‘fathom’), *cairteal* / *cairteil* (‘quarter’), *fad* (‘length’), *mìle* (‘mile’), *slat* (‘yard [measurement]’), *troigh* (‘foot [of measurement])’; cf. also ‘*fathom*’, ‘*kilometre*’

(iv) Measure

bolla (‘boll, bag’), *meud* (‘size’), *uiread* (*urrad* ‘quantity, amount’)

(v) Weight

dram (‘dram [of weight]’), *tunna* / *tonna* (‘ton’), *ùnnsa* (‘ounce’)

(vi) Currency

gròt (‘fourpence’), *not*(*a*) (‘pound [sterling]’), *peighinn* (‘penny’), *punnd* (‘pound [currency]’), *sgillinn* / *sgilling* (‘penny’), *tastan* / *tasdan* (‘shilling’)

(vii) People

duine (‘man, person’), *pearsa* (‘person’), *naodhnar* (‘nine people’), *sluagh* (‘people’)

(viii) Collective

buntàt(*a*) (‘potatoes’), *luingeas* (‘ship(s)’), *sluagh* (‘people’)

(ix) Abstract

beannachd (‘blessing’), *leum* (‘parachute jump’), *mallachd* (‘curse’), *seòrsa* (‘type’).

All nouns, with the sole exception of east Perthshire *rum*, meaning ‘room [apartment]’, can be categorised according to one of the classes above. *Rum*, which occurs as *trì rum* /trî: rúm:/ (‘three rooms / apartments’), with geminate *m*, and has the plural form *rumaichean* (pl.), contrasts with abstract noun *rùm* (‘room [space, scope]’), which can be realised with long vowel or geminate *m*,

in this dialect, i.e. /ru:m ~ rum:/ (Ó Murchú 1989: 392, 393). It is possible that the use of the singular with the concrete noun *rum* (‘room [apartment]’) has been influenced by the abstract noun *rùm* (‘room [space, scope]’).

5 Conclusion

Progress made to date on Faclair na Gàidhlig and DASG / Corpas na Gàidhlig illustrates the clear benefits which can accrue from new approaches and interdisciplinary collaboration in projects of national importance. Both projects have set the foundations for many ongoing and future projects, the outputs of which will include the production of a dictionary of Scottish Gaelic compiled on historical principles, bringing Gaelic dictionary resources abreast of Scots and English, and the means to enable scholars to contemplate seriously the production of the first comprehensive grammar of the Scottish Gaelic language. These are two of the main desiderata in Scottish Gaelic studies but Faclair na Gàidhlig and DASG / Corpas na Gàidhlig will enable much more to be achieved in terms of derived dictionaries based on the authoritative foundation of Faclair na Gàidhlig and the extensive research opportunities which DASG will stimulate. As linked complementary resources – the language corpus and the comprehensive interpretive tool – they will explain the Gaelic language from its earliest evidence to the present proving that new approaches do indeed make sense.

References

Anderson, Wendy (ed.) (2013). *Language in Scotland: Corpus-based Studies*. Scottish Cultural Review of Language and Literature (SCROLL) 19. Amsterdam: Rodopi.

Anthony, Laurence. 2005. '*AntConc*: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom'. In *2005 IEEE International Professional Communication Conference Proceedings* (Piscataway, N.J.: IEEE, 2005), 729–37. Also available at <http://ieeexplore.ieee.org> (accessed December 2012).

Anthony, Laurence. 2012. *Laurence Anthony's Website* <http://www.antlab.sci.waseda.ac.jp/software.html> (accessed December 2012).

Armstrong, Robert. 1825. *A Gaelic Dictionary in Two Parts …*. London: J. Duncan.

Craigie, William A. and Adam Jack Aitken et al. (eds.) 1931–2001. *A Dictionary of the Older Scottish Tongue: From the Twelfth Century to the End of the Seventeenth*. London: Oxford University Press.

Dareau, Margaret G. 2012. 'Dictionary of the Older Scottish Tongue'. In Macleod and McLure (2012): 116–43.

Dictionary of the Scots Language. <http://www.dsl.ac.uk>

Dwelly, Edward. 1902–11. *A Gaelic Dictionary: specially designed for beginners and for use in schools, profusely illustrated, and contains every Gaelic word in all the dictionaries hitherto published, besides many hundreds collected from Gaelic speakers and scholars all over the world.* Herne Bay: E. Macdonald.

Forsyth, Katherine (ed.) 2008. *Studies on the Book of Deer*. Dublin: Four Courts Press.

Forsyth, Katherine, Dauvit Broun and Thomas Clancy. 2008. 'The Property Records: Text and Translation'. In Forsyth (2008): 131–44.

Gillies, William and Lorna Pike. 2012. 'From Medieval Beginnings to 1911'. In Macleod and McClure (2012): 201–35.

Grant, William, David D. Murison, et al. (eds.) 1931–76. *The Scottish National Dictionary*. Edinburgh: Scottish National Dictionary Association.

Greene, David. 1992. 'Celtic'. In Gvozdanović (1992): 497–554.

Gvozdanović, Jadranka (ed.) 1992. *Indo-European Numerals*, Trends in Linguistics, Studies and Monographs, 57. Berlin & New York: Mouton de Gruyter.

Internet Archive. <http://www.archive.org/> (accessed, December 2012).

Jackson, Kenneth. 1972. *The Gaelic Notes in the Book of Deer*. Cambridge: Cambridge University Press.

MacAlpine, Neil. 1832. *The Argyleshire Pronouncing Gaelic Dictionary, to which is prefixed a concise but most comprehensive Gaelic grammar*. Edinburgh: no publisher details.

MacBain, Alexander. 1896. *An Etymological Dictionary of the Gaelic Language*. Inverness: The Northern Counties Printing and Publishing Co.

Macdonald, Kenneth D. [1987] 1994. 'Dictionaries, Scottish Gaelic'. In Thomson ([1987] 1994): 61–63.

MacilleDhuibh, Raghnall. 2008. '*Gairm*: An Aois Òir (1)'. *Aiste: Rannsachadh air Litreachas Gàidhlig / Studies in Gaelic Literature* 2: 94–119.

Macleod, Iseabail. 2012. 'Scottish National Dictionary'. In Macleod and McLure (2012): 144–71.

Macleod, Iseabail and J. Derrick McClure (eds.) 2012. *Scotland in Definition: A History of Scottish Dictionaries*. Edinburgh: John Donald.

Macleod, John et al. 1828. *Dictionarium Scoto-Celticum: A Dictionary of the Gaelic Language; comprising an ample vocabulary of Gaelic words, as preserved in vernacular speech, manuscripts, or printed works, with their signification and various meanings in English and Latin, illustrated by suitable examples and phrases, and with etymological remarks, and vocabularies of Latin and English words with their translation into Gaelic. To which are prefixed an introduction explaining the nature, objects and sources of the work, and a compendium of Gaelic grammar. Compiled and Published under the Direction of the Highland Society of Scotland*. Edinburgh: Blackwood, London: T. Cadell.

MacLeod, Norman and Daniel Dewar. 1831. *A Dictionary of the Gaelic Language*. Glasgow: W.R. M'Phun.

MacLeòid, Iain F. 2005. *Na Klondykers*. Inbhir Nis: Clàr.

Murray, James A.H. et al. (eds.) 1884–1928. *A New English Dictionary on Historical Principles*. Oxford: The Clarendon Press.

Murray, James A.H. et al. (eds.) 1933. *The Oxford English Dictionary*, being a corrected reissue with Supplement. Oxford: The Clarendon Press.

Ó Maolalaigh, Roibeard. 2008a. 'The Property Records: Diplomatic Edition Including Accents'. In Forsyth (2008): 119–30.

Ó Maolalaigh, Roibeard. 2008b. *Scottish Gaelic in Twelve Weeks*. Edinburgh: Birlinn.

Ó Maolalaigh, Roibeard. 2008c. '*"Bochanan modhail foghlaimte"*: Tiree Gaelic, Lexicology and Glasgow's Historical Dictionary of Scottish Gaelic', Scottish Gaelic Studies 24: 473–523. Also available at <http://eprints.gla.ac.uk/4830/1/4830.pdf>

Ó Maolalaigh, Roibeard (2013). '*Corpas na Gàidhlig* and Singular Nouns with the Numerals "three" to "ten" in Scottish Gaelic'. In Anderson (2013): 113–42.

Ó Murchú, Máirtín. 1989. *East Perthshire Gaelic: Social History, Phonology, Texts and Lexicon*. Dublin: Dublin Institute for Advances Studies.

Open Corpus Workbench. <http://cwb.sourceforge.net> (accessed December 2012).

Pike, Lorna. 2008. 'Supporting the Language, Defining the Way: Gaelic Dictionaries, Past, Present and Future'. Scottish Gaelic Studies 24: 525–540.

Thomson, R. L. (ed.) 1970. *Foirm na n-Urrnuidheadh: John Carswell's Gaelic Translation of the Book of Common Order*. Edinburgh: Oliver & Boyd for the Scottish Gaelic Texts Society.