

Chapter Six

Introduction to the Digital Lexical Atlas of Scotland

John Kirk, Markus Pluschkovits, Hans Christian Breuer and
Ludwig Maximilian Breuer

University of Vienna

Abstract

Almost 70 years after original data collection, the *Digital Linguistic Atlas of Scotland* project has undertaken the digitisation of those data. The present paper introduces the main challenges of the digitisation process itself and elaborates on the benefits of that digital representation. The *Digital Lexical Atlas of Scotland* provides full scope for a lexicological and semantic reanalysis and fresh interpretation of the original *Linguistic Atlas of Scotland* material. At the same time, the new atlas provides the resources for addressing various research questions (e.g., what is Scottish about the *DLAS* data? The traditional folk vocabulary of Scots: How Germanic, how Celtic? How semantic, how pragmatic are diminutives in Scots? How far does lexical variation map on to dialect patterns of phonological variation?) and for re-appraising a number of linguistic hypotheses (e.g., concerning the Highland Line, the Scottish-English border, or the 80/20 hypothesis of A-graphs).

Keywords: cultural ethnography, geography, lexemes, lexical semantics, lexical variables, maps, ortho-phonological variants, Scots

1 Introduction

Let us begin with a simple, generalised map of the Scots-speaking area of Scotland with which we will be concerned – the sweeping shaded area in Appendix, Map 1 (from Catford 1957), which runs from Orkney southwards

Kirk, John, Markus Pluschkovits, Hans Christian Breuer and Ludwig Maximilian Breuer. 2024. 'Introduction to the Digital Lexical Atlas of Scotland'. In Christine Elswailer (ed.). *The languages of Scotland and Ulster in a global context, past and present. Selected papers from the 13th triennial Forum for Research on the Languages of Scotland and Ulster, Munich 2021*. Aberdeen: FRLSU, pp. 177–245. ISBN: 978-0-9566549-7-7.

(the Shetland Islands omitted to save space), mostly along the north-eastern coast and the arable low-lying lands in the hinterland, across the central belt, on further to the border with England, and across to Ulster. In addition, the map shows two Gaelic-speaking areas, identified by the percentage of population as Gaelic speakers, and the largely uninhabited areas of the Central and North-West Highlands, labelled for ‘Highland English’ and some Gaelic. A more differentiating map is presented by Speitel and Mather (1968) (see Appendix, Map 2). On the one hand, this map demarcates the then four core areas of Scots (labelled ‘Island Scots’, ‘Northern Scots’, ‘Central Scots’, and ‘Southern Scots’), established on phonological grounds by Grant (1931), but with a couple of modifications by Mather;¹ on the other hand, it indicates areas into which Scots spread at a later date: in Shetland and Orkney and North-east Caithness, replacing Norn; in the Black Isle, parts of Moray, parts of Argyll and Kintyre, and in Galloway replacing Gaelic; and in Ulster, where it was transplanted in the early seventeenth century. The map also sets out the Highland lines identified by Grant (1931) and by Catford (1957), which roughly coincide with the western boundary of Scots in Map 1 (but see further section 6.1 below).

By the mid-twentieth century, dialectologists and historical linguists in general had become aware of the need to capture the folk vocabulary of a nation lest it might become eroded or disappear altogether under the combined effects of compulsory education (in Scotland since 1870) and growing industrialisation and the migration of peoples to urban centres. Thus, in Scotland, in 1949, a linguistic survey for both Scots and Gaelic came to be established, culminating in the publication, for Scots, of an atlas in two

¹ In Grant’s (1931) model, Ulster Scots is seen as a sub-dialect of West Central Scots. However, on grounds of contact, in this case with Irish, it can be claimed more plausibly as the fifth regional dialect of Scots (cf. Kirk 2011).

volumes for the lexical material (Mather and Speitel 1975 and 1977) and a third volume for the phonological material (Mather and Speitel 1985).²

At the time of publication, the lexical volumes received several substantial reviews, notably by McClure (1975), Macaulay (1977 and 1979), and Murison (1978). In due course, some of those data became treated in a few descriptions of Scots, notably Macaulay (1985), Tulloch (1997: 427–428), Macafee (2003: 59–61), and Millar (2007: 79–83 and 2018: 123–127). However, despite aspirations at the time of the launch of the atlas volumes, no comprehensive study of the entirety of those materials has ever taken place, until now. Nevertheless, we can readily accord with Speitel when he comments that ‘the intention to collect dialect vocabulary in order to show its distribution over the Atlas area can be considered to have been successful’. (Speitel 1969: 51)

So, in 2019, there came to be launched, at the University of Vienna, a project to convert those two published lexical volumes of *The Linguistic Atlas of Scotland (LAS)* (hereafter referred to as *LAS-1* and *LAS-2*) (Mather and Speitel 1975 and 1977) into a digital database from which maps could be interactively created and displayed, and on the basis of which a fresh analysis and interpretation of those data could be undertaken.³ A small preliminary study was undertaken by Christian Hessle and John Kirk, which not only uncovered many of the difficulties involved but also identified a number of solutions (Hessle and Kirk 2020; Kirk and Hessle 2020). In due course, Ludwig Breuer from the Sonderforschungsbereich Deutsch in Österreich joined the *Digital Lexical Atlas of Scotland (DLAS)* project and recruited his brother Hans-Christian Breuer as the programmer and subsequently his

² Cf. https://en.wikipedia.org/wiki/Linguistic_Survey_of_Scotland [accessed: 31 July 2023]. The Gaelic material is published as Cathair Ó Dochartaigh. 1994–97. *Survey of the Gaelic Dialects of Scotland: Questionnaire Materials Collected for the Linguistic Survey of Scotland*. 5 vols. Dublin: Dublin Institute for Advanced Studies.

³ A previous attempt, based on the East Scots data, is reported in Kirk (1994a).

immediate colleague Markus Pluschkovits as a further collaborator. Consequently, the data from the atlas became inputted through OCR and thoroughly checked, particularly for scanning errors.⁴ Those data comprised the register of localities, for which coordinates had to be sought, and the entire primary data, which had to be re-arranged and re-structured within the new relational database in terms of counties, localities, informants as well as the questionnaire items and responses. Various base maps were imported, including the default open street maps, a traditional ordnance survey map, and also black and white maps to facilitate non-colour publishing.

1.1 Background

The data for the lexical component of the Scots part of the Linguistic Survey of Scotland were gathered by a written questionnaire from individuals willing to undertake the task. The concepts listed in the questionnaire cover a great range of reference to human beings, their bodies, their clothes, their characteristics, to children's games, to the natural world, including insects, beasts and farm animals, and to the land and traditional (usually manual) ways of farming the land and animal husbandry – all referring to concepts which have been in oral currency for centuries. The identification of concepts – especially those that had a large number of names around the country (such as EARWIG) – had been drawn up with the assistance of A. J. Aitken and David Murison, the leading lexicographers of Scots at the time (Aitken 1954, 2015). As Catford comments: 'In Scotland we are fortunate in that good dialect speakers are to be found nearly everywhere who are literate, intelligent, and interested in their local culture and traditions, including speech.' (Catford 1957: 113). The two questionnaires were distributed in

⁴ We are most grateful to the research assistants who carried out this work: Kürsat Mutlu, Lasse Pröbsting and Erzsebet Sallai, each of whom had previously completed a BA dissertation using the *DLAS* material.

1951 and in 1953, respectively, throughout the entirety of Scotland (including the Gaelic-speaking areas) with the assistance of primary school headmasters, who were encharged with recruiting suitable respondents to complete the questionnaires. The respondents were asked to write down their ‘usual local word(s)’ for a specific concept or referent, and to use a spelling which reflected their local pronunciation.⁵ These spellings were consequently very varied, at times quite imaginative, and extremely numerous (cf. Speitel 1969).

What the present project is concerned with is to group those ortho-phonological variants of a single response into one single spelling of that response, in short to lexemise those ortho-phonological variants. No matter how varied *dog* might be spelled, it is not a *cat*. Our concern is with distinguishing the lexical types DOGS from CATS, regardless of how *any dog* or *any cat* has been spelled in the response data.

For any lexeme, a choice of spelling has to be made. That choice is often the standardised spelling which Scots shares with English, such as *ankle* for ANKLE. In many cases, however, the choice is of a Scots word. And the spelling of that word that we have chosen is usually the headword in the *SND* (e.g., *cuit* for ANKLE, *gruip* for a GUTTER IN A BYRE), although occasionally another spelling may be chosen on grounds of sheer frequency in the present data or its greater familiarity (e.g., *draidlock* instead of *draiglack* for YOUNGEST OF A BROOD). By deferring to *SND* head word choice – itself a selection from the corpus of literary and reported observational sources available to its editors – our intention is to use a spelling which has some claim to pan-Scottishness, and from which local pronunciations may be inferred.

⁵ The high figures of nil-responses are often to be explained by the presumption that, in Gaelic-speaking areas, the respondents had no local words, only the standard words.

1.2 Aims and Objectives

Thus, a first aim of the *DLAS* database project is to **lexemise** the data, i.e., to identify the lexical types underlying the ortho-phonological responses of the primary material.⁶ Patterns of non-standardised orthography intended to indicate local pronunciations are thereby being re-interpreted for the underlying lexical form.

And to help with this task, as a second aim, we also intend to etymologise the lexemic data. To date (July 2023), the ortho-phonological responses for over half of the 90 lexical variables in *LAS-I* are already analysed, with the analyses continuing expeditiously.⁷ Despite all its benefits, a sophisticated relational database does little more than store, re-arrange and re-output the data. For especially qualitative analysis and interpretation, much of the work is being done manually, requiring not only knowledge of historical linguistics but also the history of English and its pre-history as a Germanic and Indo-European language. In addition, knowledge of folklife ethnography and cultural heritage, particularly regarding traditional manual methods of farming and animal husbandry, and of traditional folk beliefs and superstitions, are also essential requirements.

A third aim is to analyse, within the traditional vocabulary of Scots as represented by the *DLAS* data, how and how far patterns of variation are to be explained in terms of the well-established internal criteria of lexical form (lexicology) and lexical semantics. However, some patterns may only come to be explained by invoking external criteria such as those social criteria of age and gender or with reference to cultural patterns within ethnology, or to

⁶ Our use of the notion of the *lexeme* derives from the well-established *-emic* (as opposed to an *-etic*) notion not only in the social sciences but also in linguistics – cf. *phonemic* and *morphemic*, as contrastive forms or types.

⁷ These analyses are being carried out by John Kirk and Julia Buser. The present goal is to have all 90 variables of *LAS-I* lexemised and analysed by 31 December 2023.

demography or (with the data coming from an atlas project) geography.⁸ In other words, the aim of the *DLAS* project is to analyse how the description of the traditional vocabulary of Scots is informed by the dimensions of lexicology/etymology, semantics, cultural ethnography and geography. These four dimensions form the core of the analysis and interpretation of the data and, together, they have already been established and vindicated through the 45 or so variables analysed and interpreted so far.

What is more, the project aims to address a number of research questions and hypotheses such as those about the nature and distribution of Scots with regard to the Highland Line (cf. e.g., Grant 1931; Catford 1957; Speitel 1981; Withers 1984 and 1988) and the English border (cf. Glauser 1974); or about the Englishness of Scots (cf. Murison 1978; Görlach 1987) or the historical make-up and diversity of Scots vocabulary. Further research questions aim to assess the distinctiveness of particular areas or, indeed, the overall dialect structure of Scots (see further section 5.4). At the same time, the project aims to consider how the interpretation of lexemes under investigation are informed by cultural practices and traditions that underlie their use.

Finally, at its most abstract, the project aims to tackle the question whether the data and their analyses and interpretations are better characterised as linguistic geography or geolinguistics (or areal linguistics).

⁸ The maps of the preliminary study by Kirk and Hessle (2020) include a toggle over the locality dots whereby the age and gender of the informant may be ascertained. Age was divided into two categories: below and above 60 years of age, allowing a loose distinction between people still working and those retired. In a study of lexemes of ANKLE, it became clear that no significant difference between the two age groups could be deduced (Kirk and Hessle 2020: 18). As for gender, whereas three of the responses of *droich* (for YOUNGEST OF A BROOD) are male, and only one female, much more evidence would be needed to trace patterns of gender difference (Kirk and Hessle 2020: 31), perhaps echoing Speitel's (1969) acknowledgement that the social dimensions of the questionnaire had not been as successful as wished for.

1.3 Data

The data underlying the lexical volumes of *The Linguistic Atlas of Scotland*, as already mentioned, are based on a written questionnaire.⁹ The questionnaire was widely distributed throughout Scotland, including the Northern Isles, the geographical areas to the west of the Highland Line, and the seven northernmost of the nine counties of Ulster. For testing the English border, the questionnaire was also distributed in the two northernmost counties of England (Northumberland and Cumberland). In the questionnaire, respondents¹⁰ were provided with a Standard English word and then asked to give their ‘usual local word(s)’, as well as one (or more) ‘less common local word(s)’ (Speitel 1969; Mather and Speitel 1975: 11). The respondents themselves were supposed to be ‘middle aged or older’ and ‘a lifelong inhabitant’ (Mather and Speitel 1975: 14) of the area for which they gave answers. They were chosen by local primary-school headmasters, to whom the questionnaires had been sent for distribution. These informants comprise people who were not at all NORMs (non-mobile, rural, older, males), so typical of many dialectological investigations but, as already indicated, more likely to be middle-class, capable of completing the written questionnaires, many possibly being those headmasters themselves, or at any rate people with whom the headmasters were acquainted. Among them were almost equal numbers of men and women, and preliminary studies reported by Kirk and Hessle (2020) and Hessle and Kirk (2020) show how little or no difference in responses on the basis of gender there is. Comparisons by age, however, show

⁹ Cf. Dollinger (2015) for a comprehensive and historiographical account of written questionnaires, including the *LAS* questionnaire.

¹⁰ In *LAS-1* and *LAS-2* respondents are referred to as ‘informants’ – two words which will be understood here as interchangeable.

some rather large differences within individual variables.¹¹ Some informants may have been quite literate, revealing a knowledge of Scots words from Scottish literature, some coming up with far more responses than ‘their usual local word(s)’.

The complete numbers of data at our disposal are presented in Tables 1 and 2. An excerpt of the originally published data for the lexical item SPIDER is presented as in the Appendix, Text 1.

Table 1: Basic Information about the LAS localities and informants

	<i>LAS-1</i>	<i>LAS-2</i>
Informants	1,774	832
Localities	1,379	731
Counties	42	42

Table 2: Basic Information about the LAS variables and variants

	<i>LAS-1</i>	<i>LAS-2</i>	Total
lexical variables	90	80	170
orthographic variants	159,660	66,560	226,220

These data are now contained within the *Digital Lexical Atlas of Scotland* database, currently hosted at the University of Vienna. It is from those data that maps are drawn automatically (see further section 4).

¹¹ Interestingly, Speitel was more dubious about age differences, as revealed in the earlier quote from him, which went on: ‘While the intention to collect dialect vocabulary in order to show its distribution over the Atlas area can be considered to have been successful, this cannot be said of the attempt to introduce social and age dimensions.’ (Speitel 1969: 51)

2 Technical implementation of the *DLAS*

The following section will set out the aims and principles of the digitisation process of the material of the original *LAS*, as well as the technology and tools used to achieve the digitisation of the original material.

The initial step in the process of digitisation is a clear understanding of the nature of the data which is to be digitised. The purpose of the *DLAS* project is the digitisation not of the original maps of the *LAS* (e.g., of *SPIDER*, as shown in Appendix, Map 3) but of the originally published data (see section 3.1) for lexemisation, the addressing of research questions and related hypotheses, and fresh cartography. The digitisation of all the response data was therefore our first goal, rather than offering digital copies of the original *LAS*-maps.¹² As has been pointed out by Macaulay (1977: 226), the word lists contain a range of more detailed and accessible data than the maps themselves, providing us with the ideal starting point for both digitisation and subsequent analysis. For Macaulay, the maps ‘add little more to the work than to provide a convenient guide to the word lists’. And of the word lists he has justly claimed:

They (the word lists) are a treasure-house of vernacular language from that area of the English-speaking world which has probably the richest variety in vocabulary of any. Where else, for example, would you find over sixty different names for ‘earwig’, or over forty different names for ‘broken pieces of china’ (used as playthings)? These lists are a striking reminder of the liveliness of the vernacular, and the loss that takes place when the medium of communication is reduced to the so-called standard language. (Macaulay 1977: 226).

¹² Nevertheless, the *DLAS* website (<https://www.lasdb.dioe.at>) does make it possible to display the original *LAS* response data. In the Explore menu, by choosing a lexical variable, there appears a list of the ortho-phonological responses, any one of which may be selected, for display as a point map.

Reviewing the data offered by the *LAS*, it becomes clear that the main entities for digitisation are the **informants**, their **locations**, the actual answers to the postal questionnaires (what we term the ‘**variants**’) as well as the individual questions contained within the postal questionnaires (what we term the ‘**variables**’) (but see further section 3.1). To achieve this, we set up a PostgreSQL database with several online front-ends for data entry developed in cooperation with the Project Part 11 of the Special Research Programme ‘Deutsch in Österreich’ (‘German in Austria’). All these tools are developed on an open-source basis, and the code can be found in the repositories of the German in Austria project (cf. *DiÖ* 2022). As mentioned above, an initial step was the import and manual OCR error-correction of localities and, based on these, the informants of the original *LAS* material. In order to geo-reference individual localities with as little manual labour as possible, the Nominatim-API in combination with OpenStreetMap (Nominatim 2022; OpenStreetMap 2022) were used. Localities not found via the API were manually added. As some locality names had changed since the data collection of the original atlas material (e.g., Peatknowe, Banff, which is no longer found on current maps but can still be found on an Ordnance Survey map of the 1920s), a further resource which proved invaluable were the geo-referenced historical maps provided by the National Library of Scotland (National Library of Scotland 2022), one of which has also been made available as a base-map for the *DLAS* (see Appendix, Map 4).¹³ To facilitate a fast workflow, a mapping interface was created, which allowed our student assistants to locate individual places and their coordinates, which could not be found automatically via a click on a map, and enter manually the coordinates found elsewhere. For localities found automatically, a manual check-up was done, where the locality could be marked as ‘correct’ via the interface.

¹³ Unfortunately, the National Library of Scotland Historical Maps API requires a subscription as of March 2022, so that this base map has been removed for now.

After localities – and thereby, the informants – were established in the database, the next step was the import of the actual data of the published atlas. In order to automate this process as much as possible, a customised OCR (i.e., Optical Character Recognition) tool was developed, which takes the formatting of the original *LAS* data into account. The standardised formatting of the data in the original *LAS* proved to be an asset in the digitisation process, as can be seen in Figure 1.

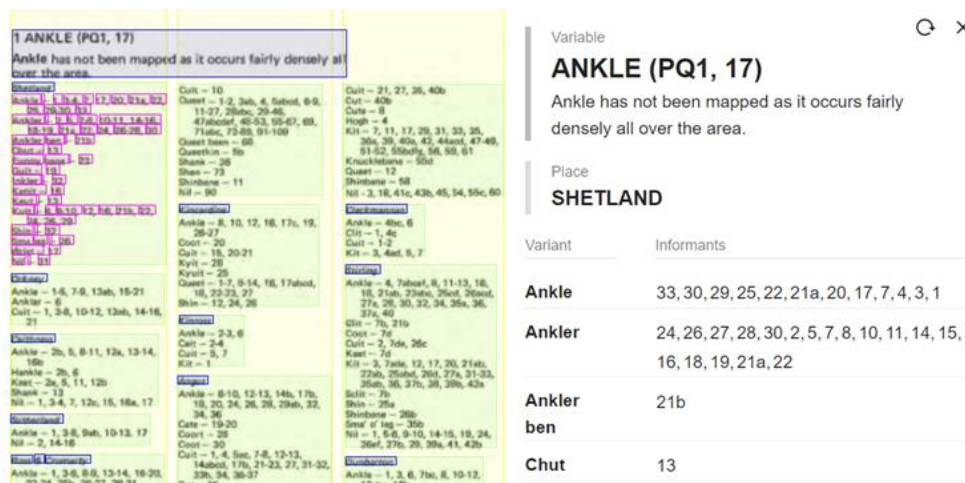


Figure 1: From Published Data, via OCR, to Digitised Data

The left-hand side of Figure 1 provides a segment of the scan of the original *LAS* material (the layout corresponds to the data shown in Appendix, Text 1) with the parsing of our OCR tool as an overlay. The blue box identifies the text within it as the headline of a data segment, which corresponds to the variable to which the data belongs. The smaller blue boxes signify the individual county names, and the purple overlay the informants by number and their responses (which correspond to the informants in the database and the variants, respectively). The right-hand side of Figure 1 shows how the information has been parsed by the OCR tool – which characters were identified as the variable, the respective informant number per county, and the individual responses. An additional third window in the middle of the tool

(omitted here for readability) provides a zoom-in on the dataset being worked on, to allow for mistakes in the OCR-process to be easily checked (typical mistakes would be lowercase i/capital I/lowercase l/the number 1, u/v, m/rn, c/o, 3/E, etc.). Mistaken characters can be corrected in the tool itself and, once checked, can be saved to the database proper.

This workflow is enabled by the relational database structure. As mentioned above, the main entities needed for the digitisation of the *LAS* data are the **informant/locality**, the **variable** (that is, the question in the postal questionnaire), and the response provided by a respondent (i.e., the **variant**). Saving a **variant** provided by an informant to a **variable** allows for flexibility with regard to the actual data mapping process, and for the dynamic creation of maps quickly and easily, with parameters far beyond the original *LAS* maps. The mapping process is further discussed in section 4.

3 Theoretical considerations

3.1 Basic semantic categorisations

Although this is primarily an empirical study which will draw inferences from distributional patterns in a large body of data, the approach is informed by a number of basic semantic concepts. The basic semantic approach in terms of which *DLAS* data is conceived is that of **onomasiology**.¹⁴ Onomasiology, deriving from Greek ὀνομάζω *onomāzo* ‘to name’, deals with the study of the names which are given to known concepts or referents and poses the question ‘what names do you give X’ or ‘what do you call X’? For the *LAS* questionnaire, which elicited the data, it was the original questionnaire item which identified the referent or concept for which responses were sought. If, for concept X, someone uses the word or expression Y, an implicit challenge

¹⁴ Some of the responses will, of course, lend themselves to semasiological treatment as well, as the same lexeme will have different meanings for different people in different parts of the country.

in the interpretation of the present data is to unravel why someone has responded to a particular concept X with the word or expression Y.

The data for the *LAS* were collected by means of a written questionnaire. A list of concepts was given (such as ANKLE, SPLINTER, LEFT-HANDED – see the many examples in the maps in the Appendix), and respondents were asked not only to write down their responses but to do so using non-standard orthography to indicate their pronunciation. Following conventions we have adopted, as mentioned in section 2 above, the original *LAS* concepts are regarded as **VARIABLES**, more specifically as **HYPERNYMS** or **LEXICAL VARIABLES**,¹⁵ and the **ortho-phonological responses**, as also mentioned in section 2, as **variants**, more precisely as **hyponyms** or **lexical variants**. These last may also be regarded as synonyms,¹⁶ but not all are strictly equivalent in meaning, with many differing in connotation and status. For our purposes, an intermediate level is introduced – the **lexical type** (i.e., the grouping of related **ortho-phonological variants** together) or **lexemes**.¹⁷ The first part of this project amounts to the lexemisation of the original ortho-phonological responses. What is being mapped are **lexemes** (but see further below). And they fall into three types: **major** (usually occurrences with relatively high frequencies of occurrence), **minor** (usually with relatively low frequencies of occurrence), and **oncurs** (*hapax legomena*, i.e., single occurrences of a lexeme). Major and minor response lexemes are thus relative both to the overall amount of responses for a variable as well as the number and distribution of responses to that variable.

¹⁵ We have adopted the convention of capitalising **HYPERNYMS** or **LEXICAL VARIABLES**.

¹⁶ Indeed, Speitel (1969: 51) regards them as ‘synonyms’.

¹⁷ For *lexemes* as well as *lexical variants*, italics are being used.

3.2 Lexemisation

The task of lexemisation is being informed by reference to numerous historical dictionaries. For Scots, there are the two monumental historical dictionaries: *A Dictionary of the Older Scottish Tongue (DOST)*, and *The Scottish National Dictionary (SND)*, now online together as the *Dictionaries of the Scots Language* (<https://dsl.ac.uk>), and condensed into the *Concise Scots Dictionary* (now revised in a superb second edition, 2017) (*CSD2*). Also relevant is *The Scots Thesaurus (ST)* (Macleod et al. 1990), based on these dictionaries, with its onomasiological approach. For Ulster, there is the *Concise Ulster Dictionary (CUD)*, edited by Caroline Macafee, one of the editors of *The Scots Thesaurus*. We also make reference to specialised dictionaries such as the *Dictionary of Scottish Building* (Pride 1996), which was edited by Iseabail Macleod and Pauline Cairns, and *Scottish Wildlife* (Robinson 2008), or regional dictionaries such as *Shetland words: A Dictionary of the Shetland Dialect* (Christie-Johnston and Christie-Johnston 2014), *The Orkney Dictionary* (Flaws and Lamb 2005) and *A Galloway Glossary* (Riach 1988) and *In My Ain Words: An East Neuk Vocabulary* (Murray 1982).

For online resources for English, there is the magnificent and recurrently updated *Oxford English Dictionary*, Third Edition (*OED3*), only available online (<https://www.oed.com>) and now with an excellent new interface; the invaluable *Online Etymology Dictionary* (<https://www.etymonline.com>) (Harper 2001–), which is based on numerous previous etymological works and recurrently updated; and the monumental *English Dialect Dictionary Online 4.0* (<https://eddonline4-proj.uibk.ac.at/edd/>) (Markus 2023), which is based on Joseph Wright's pioneering *English Dialect Dictionary* (1898–1905), and also, as version 4.0, with a magnificent new interface. As published books, there is the invaluable volume of descriptive etymologies *Word origins* (Ayto 2005), with its many illustrative folk etymologies; and three dictionaries of English Dialects:

Survey of English dialects: Dialect and grammar (Upton, Parry and Widdowson 1994), for Northumberland *A Dictionary of North East Dialect* (Griffiths 2005), and the recent *Thesaurus of English dialect and slang* (Robinson 2021).

As the basic material comprises ortho-phonological variants, studies of the phonology of Scots are being consulted as required. For lexemisation also depends on an interpretation of an item's phonological structure as indicated by the spelling. The fullest and most accessible account of dialectal variation is by Johnston (1997), although there are several older studies of traditional dialects, many cited in the *SND* and listed in Speitel and Mather (1968: Zu Karte 2). The phonological structure of words (e.g., CVC) may often serve as a starting point for lexemisation. For instance, with consonants, differences of voice may belong to the *same* lexeme (e.g., *skelf/skelve* for SPLINTER) but differences of articulation may indicate *different* lexemes (e.g., *skelf/skelb* for SPLINTER, *skelb* from Scottish Gaelic and *skelf* from Middle Dutch), with some respondents giving each variant, about which separate lexemic status may be presumed (see Appendix, Map 5).¹⁸ Vowels, however, appear to differ widely, but even although many variants belong to the same lexeme, a few do not, such as *crannie/creenie* for LITTLE FINGER (*crannie* from Dutch *krann* 'a water tap', in imitation of its shape, and *creenie* from Scottish Gaelic *crìon* 'little' (see Appendix, Map 6), thus necessitating care and attention, as with *skelf-skelb* or *sparrow*-variants or *pis*-variants (for ANT) to be mentioned in the next paragraph. Occasionally, structures have become re-aligned, such as *ettercap* for *nettercap* (for SPIDER) whereby an *ettercap* became re-analysed as a *nettercap*.

¹⁸ Occasionally, usually with oncers, a spelling which indicates a stop consonant different from that of the established lexeme may amount to no more than a slip or simple phonological confusion on the part of the respondent, e.g., *sprob* for *sprug* (for SPARROW, in Orkney).

Lexemisation is carried out in the *DLAS* database online. The process starts with the carefully checked, scanned-in **ortho-phonological responses**.¹⁹ It is proving to be a cyclic process for, in the light of experience with working with an entire set of lexemes/hyponyms, occasional revisions to the choice of lexeme or its spelling become desirable. For instance, there are seven compound lexemic variants beginning with *pis-* for ANT (notably *pishack*, *pisminnie*, *pismire*, *pismither*, *pismool*, *pismulloch*, *pismugger*, see Appendix, Map 7), but determining the exact lexeme from the range of ortho-phonological variants required careful discrimination.

Let us compare the *LAS* map for ANKLE with the *DLAS* map. The *LAS* map (Appendix, Map 8) shows *cuit/kuit*, *queet*, *cate*, and *kit* as separate items, whereas *DLAS* (Appendix, Map 9) shows them to be variants of the lexeme *cuit*. Likewise, *ankler* and *anklet* are mapped as separate lexical variants in *LAS* whereas in *DLAS* they are included simply as form variants in the map of the lexeme *ankle*. However, *DLAS* includes as lexemes *cloot*, *knuckle*, *shank* and *shin*, which are not mapped at all in *LAS*.

4 Maps and cartography

A distinction is to be made between the basic physical map and the encoding by cartographic symbols superimposed on it. The database allows for various physical maps as the base map – the default map is the freely available open street map (<https://www.openstreetmap.org>), which faintly shadows the contours of hills and mountains (see most maps in the Appendix).²⁰ The

¹⁹ As a point of methodological detail, the database facilitates the exporting of lexemisations into an Excel spreadsheet for checking and the identification of any necessary revisions. In turn, it is possible to import the Excel data into concordancing software such as AntConc, for the generation of lists by frequency or alphabetically, for further checking and the identification of any necessary revisions.

²⁰ Besides the traditional Ordnance Survey map now suspended (see Appendix, Map 4), there are also black-and-white base maps, designed to facilitate publishing.

default maps have several features: they are scalable so that a particular area can be zoomed in on; and the cartography allows for display by individual **localities** (by a simple dot, or ‘icon’) or by **counties** (with a larger dot or ‘icon’ containing a pie-chart of the lexemes for that county). The ease of map generation together with the clarity of display will play an important role in the furthering of research on the *DLAS* material.²¹

The *DLAS* database will automatically generate maps as well as the cartography superimposed on each map. However, for the creation of a map in the first instance, the lexemes (and any variants) in question have to be manually identified within the database. This is particularly crucial for lexemes associated with more than one variable (e.g., *golach* for EARWIG and also for BEETLE, CENTIPEDE, and SPIDER, or various words for GUTTER which are shared for whether along a roof, along a street, or in a byre). Maps may be compiled to show all the lexemes of a lexical variable or only some of them – specifically major types, minor types, combinations of elements involved in compounds or phrases, and oncers. By far our most common practice is to map individual lexemes and any of their variants or compounds in what we call ONLY maps – see, for instance, in the Appendix, Map 7 for *pis*-variants (for ANT), or Appendix, Map 10 for *jennie*-compounds (for SPIDER). Once a map has been compiled, it can be chosen from the map menu and be displayed. Thus, to systematise, for different purposes, various constellations of lexemes have been generated manually for the different categories of maps:

- ONLY maps, which display all occurrences (regardless of frequency) of an individual lexeme (such as *mooratoog* for ANT

²¹ The menu of map choices is being made available to end-users. See our current website at <https://lasdb.dioe.at/>. Users are asked to bear in mind that this is all provisional work in progress, that new maps will be added as and when they become available, that some maps may be revised or disappear, and that other maps may, at this stage, simply be exploratory or experimental.

(cf. Appendix, Map 11a) or *dheelog* for EARWIG (cf. Appendix, Map 11b);

- ONLY maps, which display all occurrences of a small set of lexemic variants (such as all derivatives of *grosel* [from French *groseille* ‘gooseberry’]) (cf. Appendix, Map 12);
- ONCER maps, which display all occurrences of oncers for a VARIABLE, such as all oncers for COUCHGRASS (cf. Appendix, Map 13);
- MINOR maps, which display all occurrences of the minor responses for a VARIABLE, such as all minor lexemes for CHAFFINCH (cf. Appendix, Map 14)
- MAJOR maps, which display all occurrences of the major responses for a VARIABLE, such as all major lexemes for EARWIG (cf. Appendix, Map 15);
- ALL maps, which display every occurrence of all the lexemic responses for a VARIABLE (cf. Appendix, Map 9 for all ANKLE lexemes).

In the published atlas, Scotland appears as a simple black outline on to which the data are superimposed by means of contrastive patterns of cartographical shadings. In many instances, these shadings work well; for others, especially where the patterns overlap, they become merged and render the map hard to decipher, e.g., SPIDER (*LAS-1*: Map 84, reproduced as Appendix; Map 3) or YOUNGEST OF A BROOD (*LAS-1*: Map 65). As each shading pattern covers a certain area, contrary to practice elsewhere, a perimeter line is labelled an ‘isogloss’ and defined ‘a line that surrounds a geographical area’ (Speitel 1969: 52). Such a line is rather, strictly speaking, a ‘homogloss’ marking the circumference of what within amounts to ‘identical language’ and indicating ‘a coherent linguistic area’ (Kirk

1994b: 2368). Inevitably, not all occurrences of items fall within those ‘coherent’ areas, with many appearing like outliers or ‘oncers’.²²

In the *DLAS*, by contrast, all **lexemes** are displayed by symbols, leaving their areal distribution to be inferred from the display. For maps organised by **localities**, each dot represents a single **lexeme** (of which there are many maps here); for maps organised by **county**, a small pie-chart represents the set of lexemes for that **county**. The difference may be compared between the first set of *bare*-phrases for STARK-NAKED, presented by locality (cf. Appendix, Map 16a), and the second set of *bare*-phrases for STARK-NAKED presented by county (cf. Appendix, Map 16b). The dot symbols appear in contrastive colours, for ease of readability and contrast.

Moreover, both county and locality dots contain frequency information. If a dot is clicked, the county or locality is identified, and the responses and their frequency from that county or locality for the variable in question listed.

5 The traditional vocabulary of Scots: Four dimensions

For the envisaged analyses and interpretations of the type of data we have, and to help with the answering of research questions, four dimensions are proposed: **lexicology and etymology, lexical semantics, cultural-ethnography, and geography**.²³

²² For further critical discussion see Speitel (1969), Kirk (1994b) and Kirk and Hesse (2020).

²³ By contrast, Millar (2018: 98–99) finds only three of those dimensions to Scots lexis: the historical or etymological origin of a word, its geography (although puzzlingly defined in terms of ‘purity’ and ‘genuineness’), and its semantics (defined as ‘thesaurus-like’ and thus in terms of onomasiology and hyponymy across semantic fields).

5.1 Lexicology and etymology

In the approach to the data we have adopted, we have presented a conventional linguistic model which, quickly to repeat, contrasts **lexical variables (hypernyms)**, with **lexemes (lexical types / hyponyms)** and with **ortho-phonological variants**. Our main focus is on **lexemes**, i.e., word types which are etymologically contrastive with, or formatively different from, other word types, in the way that **phonemes** and **morphemes** are contrastive but may subsume a number of realisations whether phonetic variants or allomorphs. Thus a **lexeme** may comprise a number of **form variants**, such as the form variants *ankler* and *anklet* for *ankle*, or *grosel*, *groser*, *grosert* and *groset*, along with several other variant forms, all derived from French *groseille* ‘gooseberry’ (for GOOSEBERRY) (see Appendix, Map 12) or the case of many diminutives of *quey* such as *queyack*, *queyie*, *queyo* and the double-diminutive *queackie* (for HEIFER) (see Appendix, Map 17a); or re-alignments such as *ettercap* and *nettercap* (for SPIDER), as we have seen; or the reversible compounds *heifer-cow*, *heifer-quey*, *heifer-stirk* and *cow-heifer*, *quey-heifer* and *stirk-heifer* (for HEIFER) (see Appendix, Map 17b). However, if lexemes are etymologically distinct, then we keep them separate such as, as already mentioned, *skelb* from Scottish Gaelic and *skelf* from Middle Dutch for SPLINTER, or *crannie* from Dutch *krann* ‘a water tap’, in imitation of its shape, and *creenie* from Scottish Gaelic *crìon* ‘little’ (for LITTLE FINGER). Besides, the distribution of each pair somewhat reflects those origins – *creenie* (which includes the diminutive variant *creenack*) only in the North of Scotland, *crannie* throughout Central (where it has undergone metathesis as *curnie*), East and Northern Scotland as well as Northumberland. The three occurrences of the hybrid *creenie-crannie* occur only in Morayshire (see Appendix, Map 6).

In terms of frequency, as already mentioned, we distinguish major types, minor types and oncers, but what constitutes a minor type and its frequency is coming to depend a bit on the actual variable and the nature and

extent of its exponence. Our online maps display frequencies per counties and localities in the right-hand legends (which have not been reproduced in the maps here), so that it is on those figures that we rely for frequency distributions (not the total for informants, except, of course, in the case of oncers) – and, as just mentioned, frequency per county or locality on an individual basis can always be ascertained by clicking the dot. Those lexemic maps which are made up of the base form together with significant formation variants, compounds (where a base form varies between the first or second element), or which comprise variant phrases, display a total for the entire set.

Lexemes comprise a small range of formation types: a **single stem** (e.g., *cuit* for ANKLE); **derivations** (cf. Tulloch 1997: §10.3.2.3)²⁴ such as **diminutives**, which are derived from the base form, with which the same meaning is shared (such as *creepie*, *crackie*, *currie*, *cuttie*, *crookie/croakie*, *coppie* and *cracket* for THREE-LEGGED-STOOL); and also **diminutives** which are independent of the base form (such as *pinkie* [for LITTLE FINGER]), for which the base form **pink* is not a variant).²⁵ Among derivations, diminutives certainly predominate (*pace* Tulloch 1997), occurring in abundance, so that they cry out as a research question for special

²⁴ Tulloch's observation that 'derivatives, formed by adding affixes of various kinds to existing words, are probably numerically somewhat fewer in Scots dictionaries than compounds and phrases' would certainly be testable on our data (Tulloch 1997: 401).

²⁵ Tulloch notes that 'the suffix *-ie* is particularly flexible, forming adjectives from nouns (*goskie* 'luxuriant' from *gosk* 'coarse, rank grass produced by cattle droppings') and from verbs (*grippie* 'avaricious'); making nouns from verbs (*winnie* 'a marbles game in which the winner keeps his gains' and *plottie* 'a hot drink' from *plot* 'scald'); allowing one noun to give rise to another (*batchie* 'baker', from *batch of bread*, and *steamie* 'public washhouse'). As a diminutive, *-ie* has a rival in *-ock* which gives us *devilock* and *Sannock*, an alternative to *Sandie*, and also appears in *mealock* 'crumb' derived from *meal* 'crumble'. *Mealock* gives rise in its turn to a double diminutive form *meelackie*, recorded by Jamieson in 1808 and still known in the north-east in the 1960s. [...] Some of the words formed with *-ie* or *-y* are themselves productive with proliferating meanings and compounds.' (Tulloch 1997: 401)

analysis with regard to their form and semantic function – not only in terms of the variability of their semantics, but also in terms of their pragmatic connotations of endearment and hypocorism.²⁶

Further types of lexemes are **shortenings**, such as *cap* (< *ettercap* for SPIDER); **reduplicatives** (such as *creepy-crawlie* for SPIDER, *pirlie-wirlie* for LITTLE-FINGER, *fuzzy-wuzzy* for CATERPILLAR, and many more) (cf. Tulloch 1997: §10.3.2.4); **compounds** (cf. Tulloch 1997: §10.3.2.1),²⁷ of which there are also a great many (such as *jennie-long-legs* and *spinnin'-jennie* for SPIDER or *tatty-bogle* for SCARECROW) or, as we have just seen, alternating compounds for hyponyms for HEIFER (see Appendix, Map 17a);²⁸ and **phrases**, many of which are descriptive and variable (such as *Jennie-(with-the)-hundred-legs* for CATERPILLAR, *dead-man's-bluebell* for FOXGLOVE, *one-two-year-old-heifer* for HEIFER, and *(as) bare (as a) birkie* for STARK NAKED) (see Appendix, Maps 16a and 16b); or variants of the collocation 'being born naked' (such as *as-bare-as-the-day-you-were-born*, *as-bare-as-when-he-was-born*, *as-naked-as-they-were-born*, *as-naked-when-born*, *as-the-night-he-was-born*, *as-you're-born*, *born-naked*, *like-the-hour-he-was-born*, and *same-as-you-were-born*) (see Appendix, Map 18).²⁹

Some lexemes are **onomatopoeic** (such as *maw*, *pleengie* and *screecher* for SEAGULL) or in some sense **imitative** of the referent in question (surprisingly not a category in Tulloch 1997) such as *pis*-compounds

²⁶ Of which, a preliminary study is Kirk and Pluschkovits (2023).

²⁷ Tulloch sums up his treatment of compounds by remarking rather insightfully: 'Many, probably the majority, of these compounds denote everyday things, activities, emotions and attitudes applying in a broad sphere of life, but certain specific phrases also show the language responding to particular changes in society.' (Tulloch: 1997: 399)

²⁸ In compounds, each element may vary position, as shown in the *jennie* compounds for SPIDER such as *jennie-spinner* and *spinnin'-jennie* (cf. Appendix, Map 10) or *quey*-compounds above.

²⁹ In lexemes which comprise more than one word, hyphens are used to make the lexeme discrete.

for ANT or *blob* for BLISTER, or to *jab*, *job* and *prog* for TO PRICK. The pointed or curly shape or small size of the FIR-CONE is indicated in the major types *peeries/peedies*, *burrs* and *cockabendies*, and in the minor types *apples*, *pirns*, *dottles*, *conkers* and *curlies*. Imitative creations include *cleuked/cleuchit* and *jouked* (for LEFT-HANDED). An interesting example is the response for A MAN'S UNDERVEST *peesweep* 'a miner's singlet, usually of blue-grey flannel, like a lapwing's wings', originating onomatopoeically from the bird's cry (*SND*). Such imitative and onomatopoeic forms are most likely to be coinages in Scotland.

The data for each of those lexical subcategories thus lend themselves to aggregation and investigation as group types (e.g., independent derivatives or imitatives), providing the scope for fresh analyses and interpretations, and the possibility of new distributional maps.

Accounting for **etymology** is being carried out largely in a two-stage process: identifying the **origin** of a word, which might ultimately be Proto-Indo-European, and the **donor source** of the word into Scots – the source from which the word was traditionally borrowed or evolutionarily carried over into Scots. Some words are created within English (such as *heifer* itself [from OE *hæg* 'the pen' and *fore* 'the one who moves'] whence *heifer* 'the one that moved in the pen') (cf. Lieberman 2021) or Scots (such as the double-diminutive *queyackie* [spelled in the data as *queykie*] for HEIFER). Although traditionally called **loanwords** (cf. Tulloch 1997: §10.3.1), many could be categorised as **apports**, whereby speakers of other languages immigrating to Scotland bring their own words with them and, being unwilling to give the word up, carry them over into the Scots they come to learn (cf. Kirk 2023 for an account of Irish words in English as apports, not loanwords). Thus, for instance, many words from Flemish (Middle Dutch) (cf. Aitken 1954, 2015; Murison 1971) almost certainly became absorbed into Scots in this way. As Scots evolved out of a merger of Northumbrian Old English and the northern Anglo-Scandinavian dialect of Early Middle

English, where bilingualism was almost certainly widespread, many words of Old English or from Old Norse simply evolved or ‘apported’ into Scots without ever having been ‘borrowed’ in a literal sense. It is through social and demographic history that the presence of some lexemes in Scots is being explained.

5.2 Lexical semantics

As already mentioned, the basic semantic approach informing the data is that of **onomasiology**: ‘what do you call the concept X’? Hence, what we are analysing and interpreting are those unadulterated hyponymic responses which respondents wrote down in the elicitative questionnaire, presumably in good faith, as their ‘usual local words’ (Mather and Speitel 1975: 11) for whatever particular concept X. Many of those responses can only be interpreted semantically, and semantic information and categorisations, of the type about to be described in this section, is, of course, being taken into account in the task of etymologising.

Some lexemes are straightforwardly **denotative** (‘they call a spade a spade’); but others comprise four identifiable sub-types. For a start, there are responses which amount to metonymic **transfers** – for instance, certain responses for ANKLE, such as *elbow*, *funny-bane*, *shin* or *wrist*, which are usually associated with other parts of the body, become transferred to the ankle; or else *clout*, *paster* and *tetlock*, which are usually associated with animals, especially horses, become transferred to humans. Occasionally, words for one insect were given for another (such as *golach* for SPIDER when its real meaning is EARWIG) or words for one type of GUTTER are associated with another.

Secondly, there are **descriptions** – sometimes analytic descriptions – of the phenomena in question in terms of fairly basic present-day English, some possibly intentionally humorous, such as *a bit o’ stick* for SPLINTER,

or a *farmer's friend* for SCARECROW, or the gestural use of *beggin' mare* for LITTLE FINGER.

And, thirdly, there are **metaphorisations** often involving a narrowing or widening of meaning, such as *crannie* for 'little finger' formed from *cran* 'a tap' (as mentioned, from Dutch *krann* 'a water tap', in imitation of its shape), *fired feet* or *fired hands* reflecting the cause from burning for BLISTER, or *bottle necked* or *bottle shouldered* for ROUND-SHOULDERED, whereby *bottle* is 'a rounded piece of timber running along the ridge of a roof over which a covering of lead or zinc is fixed' (SND).

Despite these identifiable and classifiable sub-types, further semantic issues are raised by the data. Another is **connotations**: for instance, the lexemes for SCARECROW seem to occupy two main semantic fields: one referring to 'a dummie man, something inhuman disguised as if it were human' (e.g., *mawkin*, *guy*, *effigy*, *dummie-man* and *wooden-man*), the other referring to a 'ghost' or 'effigy' (such as *bogie*, *bogle*, *faerie* or *worricow*). Such examples of meaning transfer raise the issue of how far **connotations** are actually **metaphorisations** – or actually, *a fortiori*, how far **metaphorisations** are simply polysemy and thus amount to de-metaphorisation. In those respects, too, there is no doubt that there will be scope for fresh assessments of those aggregated category types, but exactly how many there will be is hard to specify at this stage. Further semantic categories might include the **allusiveness** of responses such as *Benjamin* to the Bible and *Antony* to the patron Saint for YOUNGEST OF A BROOD, and for SPIDER to *Robert-Bruce*, of legendary folklore fame (cf. Kirk and Hesse 2020).

5.3 Cultural ethnography

The third dimension inherent in our data is that of cultural and ethnographical practices, in terms of which much lexical and semantic analysis will be made. In many cases, only by taking cultural and ethnographical factors into

consideration is it possible to explain why a word has come to have a certain meaning. For manifold traditional and cultural practices require a specific vocabulary. In their preliminary study, Hessle and Kirk already recognise that, in some cases ‘cultural practices or patterns of belief [...] will help identify cases of semantic transfer’ (Hessle and Kirk 2020: 6).

SCARECROWS may serve as an example of a lexical connection with a cultural practice. For SCARECROWS to exist, for instance, the cultural habits need to include the cultivation of fields for crops such as barley (cf. *barley-bogle*) or potatoes (cf. *tattie-bogle*). Moreover, cultural-ethnographical practices involve beliefs and superstitions – of which the *scarecrow* is again a good example, as a *bogey* or *bogle*, it is often believed to be a ‘ghost, spectre’.³⁰

The vocabulary of Scots is no exception. Especially among country folk, beliefs in witches, fairies, other ‘wee folk’ and the devil as well as their powers has been a prevalent characteristic of traditional life. For instance, the interpretation of the very name *foxglove* itself engenders folkloristic arguments. Is it *fox* not because of the animal or ‘the mute past of English herb-lore’ in which the name is now lost (Etymonline) but because of phonological reduction from *folks*? Which, in the context, is itself quite plausibly a reduction from the *good folks’ glove* in reference to the wee folk or to elves? (cf. Lieberman 2010). In turn, *glove* hyponyms include *bells*, *fingers* and *thimbles* (< (small) *thumbs*?) (combined in the German *Fingerhut*) probably because of their small shape. Thus, in our data, there are a great many references to *bad man’s/men’s fingers/thimbles*, *blind men’s*

³⁰ Consider the following poem, ‘The Bogle’, by W.D. Cocker, which the first-named author learned at primary school in Scotland:

There’s a bogle by the bour-tree at the lang loan heid,
I canna thole the thocht o’ him, he fills ma he’rt wi’ dreid;
He skirls like a hoolet, an’ he rattles a’ his banes,
An’ gi’es himsel’ an unco fash to fricht wee weans.

fingers/bells, dead man's bells/thimbles, devil's fingers/thimbles, fairy fingers/thimbles, lady's fingers/thimbles, wee-folk's thimbles, and witches-thimbles, each of which can only be explained through folk beliefs and superstitions, for discussion elsewhere.

Beliefs and superstitions are to be associated with lexemes in many of the main lexical **fields** in our data: personal, domestic, children's games, natural life, and farming.³¹ Linguistic data can be cross-referenced with real-life categories of 'cultural practices or patterns of belief, [which] will help identify cases of semantic transfer as shown in the use of horse words for human ankles and, in turn, for implements used in the sport of curling' (Hessle and Kirk 2020: 6).

Moreover, some cultural practices are shared far outwith Scotland. The idea of a scarecrow as a *bogey* 'ghost, spectre' or a *dummie man* can be linked to the Swiss tradition of the *Böögg*. Whereas the British tradition of *bogey(man)* (> Scots *bogle*) appears to derive, via ME *bugge*, from Welsh *bwg* 'denoting a kind of supernatural being, of uncertain origin' (*OED3*, s.v. *bug*, n.1), comparison with Swiss-German *Böögg* is striking both linguistically (with its ultimate origins being thought by *OED3* quite plausibly to be Germanic) and culturally, through the Sechseläute festival in Zürich, Switzerland, at the beginning of Spring, in which a large dummy-human statue is burnt. This example of *bogey/Böögg* thus serves to show that across Europe there are significant common traditional and cultural practices which share older etymological cognates. The use of Biblical or Saint's names (such as *Benjamin* or *Antony*) for the YOUNGEST OF A BROOD is also shown to be cross-lingual and pan-European (cf. Kirk and Hessle 2020: 27–34).

³¹ A key reference is the comprehensive 14-volume ethnographical *Scottish Life and Society: A Compendium of Scottish Ethnology* series (2000–2011). Authored works by Fenton are also relevant, such as *Scottish Country Life* (Fenton 1976), *Wirlds an' Wark 'e Seasons Roon* (Fenton 1987), and his article 'Lexicography and Historical Interpretation' (Fenton 1974).

What is more, Germanic folklore can offer an explanation for the use of *feathers* for SNOWFLAKE, often as the second element in a compound with a preceding attribution such as *snow feather*, *white feather*, *goose feather*, *Auld wife's feather*, and *Jenny's pluckin' her feathers* (see Appendix, Map 19). The semantic connection of a white feather as something quite light and soft coincides with the attributes of a snowflake. What comes to mind, however, from a cultural – and specifically Germanic – background, is the Grimms' fairy tale of *Frau Holle* (in English, *Mother Hulda* or *Old Mother Frost*), who is often depicted as an old woman shaking pillows or a blanket outside her window, thereby creating the snow. Another connection to feathers is the reference to 'a bird with feathers' through the lexeme *bunting* or *snow-bunting*, which is compared to snowflakes insofar as its plumage depicts the illusion of falling snow.³² The examples of SCARECROW, SNOWFLAKE and YOUNGEST OF A BROOD thus serve to illustrate how important culture and tradition are for the analysis of traditional lexis and semantics, even at a pan-European level.

5.4 Geography

The fourth and final dimension in our data that relates to patterns of variation is the geography of Scotland. It has undoubtedly been taken as axiomatic that geography assumes a huge part in dialectal studies. However, in terms of linguistics, the question to be asked is how geography is to be interpreted: How exactly geography is to be inferentially correlated with linguistic variation? How exactly linguistic variation is to be inferentially correlated with geography?

Maps will certainly show the geographical distribution of individual lexemes, as we have seen, for instance, for the reflexes of French *groseille*

³² See <https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-a-z/snow-bunting/> [accessed: 29 May 22].

(for GOOSEBERRY) (Appendix, Map 12) or of Scandinavian *spur* (for SPARROW) (Appendix, Map 20); by the same token, maps can also show occurrences of oncers or regionally-restricted lexemes, such as *mooratoog* (for ANT) (< ON *maurr*, ‘an ant’, + *þúfa*, ‘a mound’), only in the Northern Isles (see Appendix, Map 11a), or *dheelog* for EARWIG (> Irish *daol* ‘beetle’ + diminutive *-og*) as oncers in Co. Tyrone and Co. Fermanagh (see Appendix, Map 11b).

Although lexemes found in Shetland and Orkney may well be of an Old Norse origin, we should not immediately jump to the conclusion that Old Norse-derived words occur only in Shetland and Orkney, as the reflexes of some words occur throughout Scotland (cf. Aitken 1954, 2015; Murison 1979; Smith 1994; Tulloch 1997), such as the many Scandinavian-derived forms of *spur* for SPARROW, the ‘*passer domesticus*’. Map 20 shows no fewer than 1104 locality occurrences of *spur*-derived reflexes: *spur* itself; the diminutive *spurack*; from which the reduced form *spurg*, its diminutive *spurgie*; the diminutive *spurdie* (with an epenthetic *d* as *spur+d+ie*); and the presumably diminutive *spurish* (SND). Then from *spurg* there are the *r*-less forms *spug*, from which there are several developments: the from *spuger*, the diminutive *spuggie*; the palatalisation of the *-g* to *-dg* with the addition of the hypocoristic suffix *-er*, whence *spadger* (also *tweed-spadger*); and the dentalisation of the stop consonant to a *-d* whence *spud*, and devoiced *-t*, whence *sput*, and the diminutives *spatie* and possibly *sprottie*. Lastly, the map shows those forms which have undergone metathesis to *sprug* and *spruggie*. The oncer *sprob* is probably a corruption of *sprug*.

By the same token, although some Celtic loanwords are restricted to contact areas, such as *cuddock*, ‘a young cow, from a year or eighteen months to two years old’ (*DOST*), which occurs as a Celtic oncer in Wigtownshire, others are not and are widespread throughout Scotland (such as *golach* for EARWIG from Gaelic *gobhlag* ‘an earwig’, ‘a fork-shaped stick’) (see

Appendix, Map 15).³³ Nevertheless, because of separate historical developments, it appears to be mainly English words that occur to the west of the Highland Line, as shown by our maps.

Moreover, the traditional dialect areas of Scots (Insular, Northern, Central, Southern and Ulster) were originally established on phonological grounds (Murray 1873; *SND* Introduction 1929; Speitel and Mather 1968) (see Appendix, Map 2). More recently, by using the *LAS* as evidence, it is claimed by Tulloch (1997) that ‘while (these phonological) boundaries work reasonably well for some sounds, (they) are less satisfactory for lexical distribution’. He continues:

Words cross boundaries, and even a quick glance at *The Linguistic Atlas of Scotland* will show that the distribution of words, where there is variation within Scotland, rarely falls neatly within the boundaries of Northern, East Central, West Central, South-west Central and Southern Scots. (Tulloch 1997: 427)

Nonetheless, Tulloch concedes: ‘As a way of describing the distribution of vocabulary, the boundaries work best if taken in their broadest categories, Insular, Northern, Central, Southern and Ulster.’ He urges: ‘It is best to consider individual referents one by one, as the *LAS* does, and then one becomes quickly aware of the diversity of patterns of distribution.’ (Tulloch 1997: 427). This is exactly the approach of the *DLAS* being taken, which is hopefully going some way towards ‘classify(ing) the vocabularies of the Modern Scots dialects’ envisaged by Macafee (in Aitken 1954, 2015: footnote 44).

Another approach to geography is to consider individual dialects – whether areas, counties, or individual localities. ‘The subject can also be

³³ On Gaelic loan words, cf. Aitken (1954, 2015), McClure (1986), Pödör (1995–1996), Tulloch (1997: §10.3.1.1) and Dareau (2001). Popular treatments include McLennan (2010) and Newton (2021).

looked at by considering not individual referents, but individual dialects’, suggests Tulloch (1997: 429). He finds that ‘some of the greatest differences lie perhaps at the two ends of the Scots-speaking area, in Ulster and Insular Scots with their very different special influences.’ (Tulloch 1997: 429) Many areas of Scotland have local identities such as the various fishing communities (cf. Map 2, Millar, Barras and Bonnici (2014)) or Galloway (cf. Map 2; Riach 1977 and 1988; Pröbsting 2020), with associations stretching back to ancient British settlements, or the legendary peninsula of Fife, with its Pictish as well as medieval royal connections (cf. Murison 1978; Murray 1982).

So, if origin and present-day geography do not necessarily go hand in hand, for the explanation for patterns of variation in the present data, we certainly envisage an interpretive role for geolinguistics and specifically its basic tenet: How is linguistic variation to be interpreted in terms of geography? That is quite different from the traditional question about where geographically a certain word is to be found – the hallmark of traditional atlas projects.

This question has already received one answer: maps of dialect variation in Scotland can be compared with economic maps of Scotland. One economic map (see Appendix, Map 21) – a milestone of its kind, drafted by Catherine Snodgrass and published in 1943 – is based on the natural topography of Scotland, which divides between literally high-lying lands (over 1500 feet) and low-lying lands (under 1000 feet), whence the usual nomenclature. The data for her agricultural maps appear to have been compiled from her own knowledge and research, whereas those for coal mining and industry appear to derive from various maps and sources, especially, it seems, *A Royal Commission on the Location of Industry*; her motivation was to assist ‘the practical side of planning for a better future for Scotland’ (Snodgrass 1943: 15). ‘The SGM (Scottish Geographical Magazine) map was a reduction and simplification of a coloured map at ten

miles to one inch scale, designed as an interim base map for planning an improved post-war Scotland', according to Kenneth Maclean [n.d.]³⁴.

In Snodgrass's map, the Highlands split into three areas: the Central and North-West Highlands (Number 1A on Snodgrass's map), the Southern Uplands (Number 1B), and the isolated Hill Masses of the Lowlands (region 1C). The Lowlands split into nine areas – six to do with types of farming, and the remaining three to do with coal mining (Number 8), industry (Number 9) and fishing (Number 10). The six farming areas have their characteristics: cattle and sheep-rearing in the Far North (Number 2), cattle-rearing and beef-producing in the North-East (Number 3), mixed farming and crop production in the lower coastal districts (Number 4), sheep-rearing and mutton-production in the Tweed Valley (Number 5), dairy-farming in the Western and Central regions (Number 6), and finally sheep and cattle-rearing with dairying on the better land in the South West (Number 7).³⁵

When set alongside maps of the dialect structure of Scots, as evidenced, e.g., by Speitel and Mather's (1968) map (Appendix, Map 2), what Macafee (1989) finds is a remarkable visible relationship between dialect divisions and farming divisions, in a word 'continuity', about which she comments: 'there is a dialect (and culture) boundary which persists over a long period of time, because it is determined by geographical factors' (Macafee 1989: 432–434). And it is that continuity from the huge geographical as well as economic spread of farming across much of Scotland that connects with and has contributed so much to the traditional folk vocabulary which is now collected as the data underpinning the present

³⁴ Quote taken from <https://www.rsgs.org/blog/memorable-maps-new-economic-map-of-scotland>.

³⁵ As a further example, Macafee (1989: 432) cites the boundary between the Anglian and Mercian dialects of Old English, what later became the boundary between Northern and North-Midland dialects in England, for that southern boundary of Anglian coincides with the ethnographical boundary of the medieval Scandinavian settlement.

project. As more variables from agriculture and animal husbandry become analysed, the more our eye will be on the correlation of words with those equally traditional and long-surviving land-use practices. It is thus essential for geography to be included as one of the dimensions of analysis and interpretation.

And so, the association of those four dimensions of lexical form, lexical semantics, cultural-ethnographic practices, and geography strike us as both necessary but also sufficient for enabling as thorough and comprehensive an analysis and interpretation of the original *LAS* data as possible.

6 Research questions for testing

The *DLAS* database allows us to explore a number of research questions and test certain related hypotheses.

6.1 The Highland Line

The notion of the ‘Highland Line’ initially began as a somewhat ‘vague’ (Speitel 1981) physical boundary between the Highlands and the Lowlands which, in due course, came to represent the boundary between Gaelic and Scots. The attempts by Grant (1931) and Catford (1957) to identify its exact geographical position are shown in Map 2, with only a few differences: whether, for instance, the Kintyre peninsula and the Isle of Arran fall within the line or not.³⁶ A central question is how far does it separate the mountainous topography of the Highlands from the more arable farmlands of the Lowlands? One attempt described by Speitel even includes Fife within the line – it all depends on what historical period is being discussed, as

³⁶ Although not depicted in Map 2, Grant’s line originally includes the Black Isle and runs all the way up the coast to include the fishing communities before connecting with its continuation in Caithness.

Dunfermline was once the capital and there was a royal palace at Falkland nearby. Nowadays, suggests Speitel (1981), the Highland Line is more likely to be thought of as the imaginary boundary between Scots dialects and the English which replaced Gaelic (see Map 1).

The first postal questionnaire of the survey which led to the *LAS* included a multiple-choice elicitation of some 14 items about pronunciation. For instance, did a person pronounce TWO as in English *two*: TOO, or rhyming with *law*: TAW, or rhyming with English *day*: TAW? The first one marked a speaker as ‘English’, the second two as ‘Scots’. By weighting certain features for Scots, Speitel (1981) calculated that the maximum possible score amounted to 26. Localities with at least 10 points, once identified, came to be linked and thereby a line was drawn – a new Highland Line, based both (conveniently for our purposes) on *LAS* localities and on the limits of what Speitel regarded as Scottish dialect-pronunciation features. The map is reproduced as Appendix, Map 22. As this line has been identified by *LAS* localities, we hope in due course that it will be added as an option to our maps, so that the distribution of lexemes on either side of the line can be seen at a glance and thereby be tested: how Scots, how English?

6.2 The Scottish-English ‘linguistic’ border

Whereas Scots shares with Northern English many words (especially those of Scandinavian origin) which transcend the Scottish-English ‘linguistic’ border, others certainly do not (cf. Glauser 1974; Aitken 1984, 2015), leading Speitel to comment: ‘I believe I am not exaggerating if I say that the Scottish-English Border is probably one of the most striking geographical linguistic divides in the English-speaking world which can be established on the basis of existing Surveys.’ (Speitel 1969: 55). Such strong indications are already inferable from Speitel and Mather’s earlier map of 62 isoglosses running across the border area (Speitel and Mather 1968: Karte 3), although it is unclear how many of those isoglosses are lexical.

Our findings so far show mixed results. Very frequent words such as *groser/groset* for GOOSEBERRY, *pinkie* for LITTLE FINGER, *cuit* for ANKLE, *humphie-backit* for ROUND SHOULDERED, *froth* for SOAP SUDS, *bogle* for SCARECROW, all of which occur throughout Scotland, also occur in the northernmost counties of England. By contrast, other frequent (or very frequent) words which also occur throughout Scotland do not occur in the northernmost counties of England. Examples would be *golach* for EARWIG, both *skelf* and *skelb* for SPLINTER, *moggans* for MITTENS and *taps* for FIR CONES. And in the two northernmost counties of England, there are many words which do not occur in Scotland, such as *cowie* for LEFT-HANDED or *ear-finger* or *fourth finger* for LITTLE FINGER or *clocker beetle* or *horse beetle* for BEETLE, and so on. With the inclusion of the two northernmost English counties in the DLAS, the border hypothesis certainly lends itself to fresh examination and testing.

6.3 The Englishness/Germanicness of Scots

For many present-day Scots-language activists, Scots is simply not English, no matter how Scots may be categorised or described. But given that English has its basis initially in Old English, Scots would appear to share that basis. As already mentioned, the Northumbrian dialect of Old English became Scandinavianised, and it was from that subsequent northern Anglo-Scandinavian dialect of Early Middle English that, through immigration northwards, Scots – written and spoken at the time – came to evolve. For Aitken ‘by far the most important single element in the Older Scots vocabulary is the Anglo-Saxon’ (Aitken 1954, 2015: 6). And for Catford: ‘The term “Scots” is used to designate a range of distinctively Scottish dialects of English [...]’ (see Maps 1 and 2) (Catford 1957: 109).

The question thus arises: How linguistically English (and *a fortiori* Germanic) does that actually make Scots? (cf. Görlach 1987) For one of the original researchers, Scots dialects were simply dialects of English (Catford

1957: 109). As Millar (2018) states quite categorically: ‘the greatest part (not necessarily the majority) of Scottish vocabulary derives directly from Old English.’ (Millar 2018: 99). From our present investigations, despite some having Scots forms with popular associations as culturally Scottish (such as *bogle* for SCARECROW), there are certainly a great many etymological Anglo-Saxonisms in our data, and in due course we will be able to test this Englishness/Germanicness hypothesis afresh. ‘At the same time, however,’ Millar goes on, ‘a considerable amount of Scots vocabulary is of North Germanic origin’, and we are certainly seeing lots of Old Norse words in our data (see discussion of *spug* below).

6.4 Issues of loss/erosion and survival

An issue recurrently raised about dialect words is their currency. As the data was collected in the early 1950s, the question of present-day currency inevitably follows. John Kirk was actually taught by two of the *LAS* respondents, one of whom was of a similar age as his mother, to whom, as well as himself, as a student, he subjected the *LAS* questionnaire. The result was that Kirk knew at that stage of his career only 25 per cent of the words provided for his birthplace town of Falkirk by the informants – mostly explicable by his lack of knowledge of the farming or animal husbandry concepts in questions (cf. Kirk 2002). The question of loss and erosion has been raised by Görlach (1987), Macafee (1994a and b) and Tulloch (1997: §10.4), but the other side of the coin is survival, and even creation, as addressed by Macafee (1991 and 1994a) and Tulloch (1997: §10.3.7). It is no surprise that many survivals relate to parts of the body or human attitudes.

6.5 Kretzschmar’s and Burkette’s A-graphs

According to Kretzschmar and Burkette (2017), in a dialect survey, 80 per cent of the data is accounted for by 20 per cent of the responses. They show this with reference to the lexical variable PARLOUR, which was elicited by

the written *LSS* questionnaire survey, but not published in the atlas. They comment:

The PQ2 returned 133 distinct response types, which account for a total of 1123 individual responses. The most common *LSS* responses were *parlour*, *sitting room*, and *best room* which together account for 66% of the total responses, a percentage that follows the spirit of the 80/20 guideline. In addition, of the 133 response types, 103 were given only once or twice.

(Kretzschmar and Burkette 2017:11)

This is shown by the A-graph which they have drawn, as in Figure 2. In short, the A-graph claim is that, in a data collection, such as that for the *DLAS*, there are a few responses which occur very frequently, and a great many responses which occur rarely. It is our impression that, from the data analysed so far, their hypothesis is, in general, plausible. However, it remains to be seen how far an 80/20 distribution will square with our MAJOR, MINOR and ONCER types, and also how far it will assist with wider analyses and explanations of variation in the frequency of lexemes of any given variable.

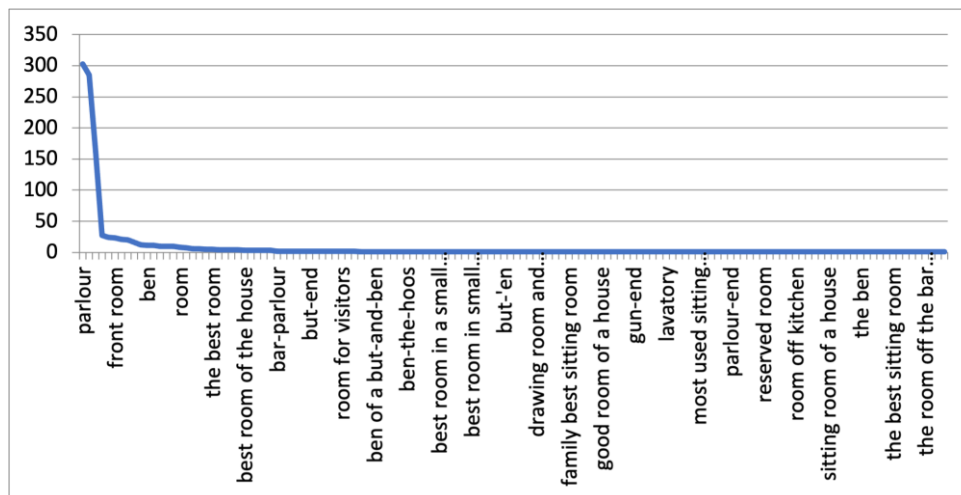


Figure 2: A-graph of lexemes for PARLOUR (from Kretzschmar and Burkette 2017)

6.6 Regional hypotheses

The *DLAS* is enabling investigations of many local areas in Scotland, such as the Northern Isles, the North-East, Galloway, or any county for that matter. On the basis of the *LAS* maps, Murison (1978) considers Fife to be a dialect island, with the Forth and Tay rivers forming southern and northern boundaries. When all the data are taken into account and lexemised, however, the rivers do not appear so watertight (cf. Kirk 1994a). Riach (1977) has investigated the historical area of Galloway, claiming, on the basis of his own fieldwork, its vocabulary to be of 50 per cent Celtic origin. By contrast, the *DLAS* data for Galloway has been analysed by Pröbsting (2020), who finds the figure to be less than 5 per cent, with the vast majority of lexemes being Anglo-Saxonisms. Many similar regional investigations could now be undertaken afresh.

A particular regional case-study case is almost certainly Ulster Scots, which was taken there by settlers in the late sixteenth and early seventeenth-century, is better considered as the fifth regional dialect of Scots on the grounds of contact with Irish (Kirk 2011) than as a sub-dialect of West-Mid Scots (according to the *SND*), whence much of it came. The *DLAS* contains data for seven counties in the North of Ireland: County Donegal plus the six counties making up present-day Northern Ireland. A dictionary of present-day Ulster Scots – *The Hamely Tongue: A personal record of Ulster Scots in Co. Antrim* (Fenton 2014) – contains numerous loans from Irish; and in the *Concise Ulster Dictionary* (Macafee 1996) the per cent of words from Scots has been calculated to 62% (that from Irish some 12%) (cf. Kirk 1999). The ortho-phonological responses of some 45 variables of the original *LAS* data are analysed and manually mapped by Zwickl (1996) for Ulster, who produces some interesting isoglosses around particular ortho-phonological forms, e.g., demarcating *ankler* from *anklet* variants (for ANKLE) (Karte 4.1), *pismool* from *pismire* variants (for ANT) (Karte 4.2), and *crowl*, *droylie*, *droich*, from *jorie* (for YOUNGEST) (Karte 4.45). She finds that English,

Irish and Scots words occur throughout Ulster, and that even relic area clusterings of Scots words in Antrim (Karte 5.5), Irish words in the North-west (Karte 5.6) or English words in the South-East (Karte 5.7) are not exclusive.

6.7 Linguistic geography vs. geo(graphical) linguistics/areal linguistics

By plotting symbols which correspond to data on physical maps, we are able to infer which words occur where, and what the geographical pattern of those words is. This has traditionally been considered linguistic geography, as the map enables us to make inferences about the data. Without the map, such insights could not be drawn. However, to re-phrase Kirk (1994b), in recent years, stimulated by sociolinguistics whereby linguistic variation could be correlated with social variables, attempts have been made to explain variation in terms of the physical landscape, as if it is the geography that is causing or is in some way determining the linguistic variation. The qualitative approach of historically oriented, paradigm-based, item-centred, ‘traditional’ dialectology has now become complemented, a hundred years later, with the quantitative approach of sociolinguistics, variation theory, and now geolinguistics. Geolinguistics seeks to relate linguistic variables and their underlying systems not only with the implicit social characteristics of the speaker’s identity (the approach of sociolinguistics) but also with any geographical factors (in the widest possible sense of ‘geographical’) which might contribute to that distribution and correlation of identity. While more complex than internally focused historical or comparative philology, the geolinguistic approach is also more realistic, for it correlates analyses of data with qualitative and quantitative analyses of external conditioning factors.

Older and newer approaches both recognise that the linguistic map is a central research tool and a principal goal. The maps that we are generating lie at the heart of our methodology. According to one German dialectologist, ‘ever since dialectology has been practiced as dialect geography, the map and

the collection of maps in the form of the linguistic atlas has become the most important means of presentation and research' (Löffler 1980: 70); and according to a leading U.S. innovator:

As a graphic resource in dialect research, computer programs produce basic charts for work in linguistic geography – the discipline that takes the linguistic atlas (a gathering of descriptive maps) as its traditional emblem and ultimate goal. (Pederson 1988: 165)

In that tradition, as Catford writes:

This is not the place to discuss the interpretation (linguistic, historical or sociological) of the distribution maps we have made, but it may be said that word geography is at least as fruitful of problems, and of their solutions, in Scotland as in other countries. (Catford 1957: 117)

7 Applications of *DLAS*

We envisage that, in due course, the *DLAS* will have many applications. As indicated above, it is already acting as a resource for many lexicological research questions and the testing of the various hypotheses just outlined. Due to popular interest, the data lend themselves not only to scientific enquiry, but to more popular journalistic treatment, for sharing the resources and findings with as broad a public as possible. Connections between vocabulary, cultural ethnography and folklore always have popular appeal.

At the same time, the *DLAS* has huge potential as a data resource for numerous research theses, which might come to take a holistic view of the data as indicative of the Scots lexicon, and which might address quite specific topics such as loanwords, diminutives, or metaphorisations. As mentioned, the *DLAS* offers great potential for regional studies (e.g., Shetland, Ulster, or simply any region or county). The *DLAS* lends itself for use in schools, where the material from any school locality could be instructive about the pupils'

traditional local dialect, but which could be checked out afresh by the pupils for currency, knowledge and use as well as issues of loss and survival. It would not be difficult to prepare suitable teaching materials.

More generally, the *DLAS*, by being available online, will serve as a resource for the general public not just locally but across the world for all kinds of lexical enquiries. In due course, the database might migrate from Vienna to Scotland, where, after all, it will be more fittingly at home. Finally, the data as well as the analyses and interpretations will almost certainly contribute to the perennial core debate of ‘what exactly is Scots?’.

Acknowledgements

We are hugely grateful for the permission given to John Kirk by Croom Helm, now some time ago, to digitise the maps and data of *LAS-1* and *LAS-2*; to Allison Burkette for permission to reproduce the unpublished A-graph reproduced here as Figure 2; to Caroline Macafee, who kindly let us have comments on a pre-publication draft and reminded us of her 1989 paper; to Manfred Markus for his valuable comments to the final draft; to the volume editor Christine Elswailer for her perceptive comments on several drafts; and to Colin Young for his artwork for Map 2. During the preparation of this article, John Kirk has been grateful for discussion with Simone Zwickl, Colin Neilands, and especially Julia Buser, who provided the references to the Böögg and Frau Holle as well as a number of useful comments on an earlier draft, and who has been assisting with the lexemisation and interpretation of the data in preparation for her PhD thesis at the Ludwig-Maximilians-Universität München.

References

- Aitken, Adam J. 1954, 2015. ‘Sources of the vocabulary of Older Scots’. In
†A. J. Aitken, ed. Caroline Macafee, ‘Collected writings on the Scots
language’ (2015), *Scots Language Centre*.

https://d3lmsxlb5aor5x.cloudfront.net/library/document/aitken/Sources_of_the_vocabulary_of_Older_Scots.pdf [accessed: 29 July 2023].

Aitken, Adam J. 1984, 2015. ‘Scottish accents and dialects’. In †A. J. Aitken, ed. Caroline Macafee, ‘Collected writings on the Scots language’ (2015), *Scots Language Centre*.
https://d3lmsxlb5aor5x.cloudfront.net/library/document/aitken/Scottish_accents_and_dialects.pdf [accessed: 13 June 2022].

Aitken, Adam J., Angus McIntosh, and Hermann Pálsson (eds.) 1971. *Edinburgh studies in English and Scots*. London: Longman.

Asher, Ronald E., and J. M. Y. Simpson (eds.). 1994. *The encyclopedia of language and linguistics*. Vol. 5. Oxford: Pergamon Press.

Ayto, John. 2005. *Word origins: The hidden histories of English words from A to Z*. 2nd ed. London: A & C Black.

Barrow, G. W. S. (ed.) 1974. *The Scottish tradition: Essays in honour of Ronald Gordon Cant*. Edinburgh: Scottish Academic Press.

Benskin, Michael, and Michael L. Samuels (eds.) 1981. *So meny people longages and tonges: Philological essays in Scots and medieval English presented to Angus McIntosh*. Edinburgh: Middle English Dialect Project.

Catford, J. C. 1957. ‘The Linguistic Survey of Scotland’. *Orbis* 6(1): 107–121.

Christie-Johnston, Adaline, and Alastair Christie-Johnston. 2014. *Shetland words: A dictionary of the Shetland dialect*. Lerwick: the Shetland Times.

Corbett, John, J. Derrick McClure, and Jane Stuart-Smith (eds.) 2003. *The Edinburgh companion to Scots*. Edinburgh: Edinburgh University Press.

CSD2 = Concise Scots Dictionary. 2017. 2nd ed. Edinburgh: Scottish Language Dictionaries and Edinburgh University Press.

- CUD* = *Concise Ulster Dictionary*. 1996. Ed. Caroline Macafee. Oxford: Oxford University Press.
- Dareau, Margaret. 2001. 'The Scots/Gaelic interface: A new perspective'. In Kirk and Ó Baoill (2001): 237–256.
- DiÖ. 2022. *GitHub repositories of the research project 'Deutsch in Österreich' ('German in Austria')*. N.p. <https://github.com/german-in-austria/lasDB/> [accessed: 17 April 2023].
- Dollinger, Stefan. 2015. *The Written questionnaire in social dialectology: History, theory, practice*. Amsterdam and Philadelphia: Benjamins.
- DOST* = *A Dictionary of the Older Scottish Tongue*. 1931–2002. Eds. William Craigie, Adam J. Aitken, James A. C. Stevenson, and Marace Dareau. Oxford: Oxford University Press. <https://dsl.ac.uk/> [accessed: 17 April 2023].
- DSL* = *Dictionary of the Scots Language* (combining *A Dictionary of the Older Scottish Tongue* and *The Scottish National Dictionary*). <https://dsl.ac.uk/> [accessed: 17 April 2023].
- Fenton, Alexander. 1974. 'Lexicography and historical interpretation'. In Barrow (1974): 243–258.
- Fenton, Alexander S. 1976. *Scottish country life*. Edinburgh: John Donald.
- Fenton, Alexander S. 1987. *Wirds an' wark 'e seasons roon on an Aberdeenshire farm*. Edinburgh: Mercat Press.
- Fenton, Alexander S., and Donald A. MacDonald (eds.) 1994. *Studies in Gaelic and Scots: Proceedings of the Third International Conference on the Languages of Scotland*. Edinburgh: Canongate Academic.
- Fenton, Alexander S. (gen. ed.) 2000–2011. *Scottish life and society: A compendium of Scottish ethnology*. Edinburgh: Birlinn.
- Fenton, James. 2014 [1996]. *The Hamely Tongue: A personal record of Ulster Scots in Co. Antrim*. Fourth edition. N. p.: Ullans Press.
- Flaws, Margaret, and Gregor Lamb. 2005 [2001, 1996]. *The Orkney Dictionary*. Kirkwall: The Orkney Language and Culture Group.

- Glauser, Beat. 1974. *The Scottish-English linguistic border: Lexical aspects*. Bern: Francke.
- Görlach, Manfred. 1987. 'Lexical loss and survival: The case of English and Scots'. *Scottish Language* 6: 1–20.
- Grant, William. 1931. 'Introduction'. In Grant and Murison (1931–1946, vol. 1): ix–xli.
- Grant, William, and David Murison. 1931–1946. *The Scottish National Dictionary*, 10 vols. Edinburgh: Scottish National Dictionary Association.
- Griffiths, Bill. 2005 [2004]. *A Dictionary of North East dialect*. 2nd ed. Newcastle-upon-Tyne: Northumbria University Press.
- Harper, Douglas. 2001–. *Online Etymology Dictionary*. N.p. <https://www.etymonline.com/> [accessed: 17 April 2023].
- Hessle, Christian, and John M. Kirk. 2020. 'Digitising collections of historical linguistic data: The example of the *Linguistic Atlas of Scotland*'. In Los, Molineaux and Mäkinen (2020): 1–17.
- Hickey, Raymond (ed.). 2023. *Oxford handbook of Irish English*. Oxford: Oxford University Press.
- Johnston, Paul. 1997. 'Regional Variation'. In Jones (1997): 433–513.
- Jones, Charles (ed.) 1997. *The Edinburgh history of the Scots language*. Edinburgh: Edinburgh University Press.
- Kirk, John M. 1994a. 'East Central Scots: A computerised mapping package'. In Fenton and MacDonald (1994): 48–68.
- Kirk, John M. 1994b. 'Maps: Dialect and language'. In Asher and Simpson (1994): 2363–2377.
- Kirk, John M. 1999. 'The Dialect Vocabulary of Ulster'. *Cuadernos de Filología Inglesa* 8: 305–334.
- Kirk, John M. 2002. 'Measuring lexical variation and change: Scots and Hiberno-English'. Paper presented at the Lexicological Workshop, Leuven.

- Kirk, John M. 2011. 'Scotland and Northern Ireland as Scots-speaking communities'. In Kirk and Ó Baoill (2011): 193–205.
- Kirk, John M. 2023. 'The Vocabulary of Irish English'. In Hickey (2023): 179–203.
- Kirk, John M., and Christian Hesse. 2020. 'Towards a *Digital Lexical Atlas of Scotland*'. *Scottish Language* 39: 1–39.
- Kirk, John M., and Dónall Ó Baoill (eds.) 2001. *Language links: The languages of Scotland and Ireland*. Belfast: Queen's University.
- Kirk, John M., and Dónall P. Ó Baoill (eds.) 2011. *Sustaining minority language development: Northern Ireland, the Republic of Ireland, and Scotland*. Belfast: Cló Ollscoil na Banríona.
- Kirk, John M., and Markus Pluschkovits. 2023. 'The *Digitised Lexical Atlas of Scotland*: The Scotsness of Scots, and issues of diminutivity'. Paper presented at the Austrian Association for University Teachers of English, Salzburg, 5 May 2023.
- Kirk, John M., Markus Pluschkovits, Hans Christian Breuer, and Ludwig Maximilian Breuer. In preparation. *The Digital Lexical Atlas of Scotland*. University of Vienna and Austrian Academy of Sciences. <https://lasdb.dioe.at/> [accessed: 17 April 2023].
- Kirk, John M., Stewart Sanderson, and J. D. A. Widdowson (eds.) 1985. *Studies in linguistic geography: The dialects of English in Britain and Ireland*. London: Croom Helm.
- Kretzschmar William A., Jr., and Allison Burkette. 2017. 'Complex systems and British Isles survey data' (typescript).
- Lieberman, Anatoly. 2010. 'Etymologists at war with a flower: Foxglove'. <https://blog.oup.com/2010/11/foxglove/> [accessed: 28 August 2023].
- Lieberman, Anatoly. 2021. 'Still plowing with my heifer'. *OUP Blog*. <https://blog.oup.com/2021/06/still-plowing-with-my-heifer/> [accessed: 12 July 2023].

- Löffler, Heinrich. 1980. *Probleme der Dialektologie*. 2nd ed. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Los, Bettelou, Benjamin Molineaux, and Marti Mäkinen (eds.) 2020. ‘Special issue on visualisations in historical linguistics’. *Journal of Data Mining and Digital Humanities*. <https://jdmhd.episciences.org/page/special-issue-on-visualisations-in-historical-linguistics> [accessed: 17 April 2023].
- Macafee, Caroline. 1989. ‘Middle Scots dialects – extrapolating backwards.’ In McClure and Spiller (1989): 429–441.
- Macafee, Caroline. 1991. ‘Acumsinery: Is it too late to collect traditional dialect?’ *Folk Life: Journal of Ethnological Studies* 30(1): 71–77.
- Macafee, Caroline. 1994a. *Traditional dialect in the modern world: A Glasgow case study*. Frankfurt am Main: Lang.
- Macafee, Caroline. 1994b. ‘Dialect erosion with special reference to Urban Scots’. In Fenton and MacDonald (1994): 69–80.
- Macafee, Caroline. 2003. ‘Studying Scottish lexis’. In Corbett, McClure and Stuart-Smith (2003): 50–71.
- Macaulay, Ronald. 1977. ‘Review of Mather, J. Y.; Speitel, H.-H. (eds.) 1975. *The Linguistic Atlas of Scotland: Scots section*, vol. 1. (London: Croom Helm, Ltd.)’. *Language* 53(1): 224–228.
- Macaulay, Ronald. 1979. ‘Review of Mather, J. Y.; Speitel, H.-H. (eds.) 1975. *The Linguistic Atlas of Scotland: Scots section*, vol. 2. (London: Croom Helm, Ltd.)’. *Language* 55(1): 224–228.
- Macaulay, Ronald. 1985. ‘Linguistic maps: Visual aid or abstract art?’. In Kirk, Sanderson and Widdowson (1985): 172–186.
- Maclean, Kenneth. [n.d.] ‘Memorable maps – new economic map of Scotland’. <https://www.rsgs.org/blog/memorable-maps-new-economic-map-of-scotland>. [accessed: 28 July 2023].
- Markus, Manfred. 2023. *English Dialect Dictionary Online 4.0*: <https://eddonline4-proj.uibk.ac.at/edd/> [accessed: 27 September 2023].

Introduction to the Digital Lexical Atlas of Scotland

- Mather, J. Y., and Speitel, H.-H. (eds.) 1975. *The Linguistic Atlas of Scotland: Scots section*, vol. 1. London: Croom Helm.
- Mather, J. Y., and Speitel, H.-H. (eds.) 1977. *The Linguistic Atlas of Scotland: Scots section*, vol. 2. London: Croom Helm.
- McClure, J. Derrick. 1975. 'Review of Mather, James Y.; Speitel, Hans-Henning (eds.) 1975. *The Linguistic Atlas of Scotland. Scots Section. Volume 1.* (London: Croom Helm Ltd.)'. *American Speech* 51(3-4): 223-234.
- McClure, J. Derrick, and Michael Spiller (eds.) 1989. *Bryght lanternis: Essays on the language and literature of medieval and renaissance Scotland*. Aberdeen: Aberdeen University Press.
- McClure, J. Derrick. 1986. 'What Scots owes to Gaelic'. *Scottish Language* 5: 85-98
- McLennan, George. 2010. *Slogans galore! Gaelic words in English*. Glendaruel: Argyll Publishing.
- Millar, Robert McColl. 2007. *Northern and Insular Scots*. Edinburgh: Edinburgh University Press.
- Millar, Robert McColl. 2018. *Modern Scots: An analytical survey*. Edinburgh: Edinburgh University Press.
- Millar, Robert McColl, William Barras, and Lisa Marie Bonnici. 2014. *Lexical variation and attrition in the Scottish fishing communities*. Edinburgh: Edinburgh University Press.
- Murison, David. 1971. 'The Dutch element in the vocabulary of Scots'. In Aitken, McIntosh and Pálsson (1971): 159-176.
- Murison, David. 1978. 'Review of Mather, James Y.; Speitel, Hans-Henning (eds) 1975. *The Linguistic Atlas of Scotland. Scots Section. Volume 2*'. *The Scottish Review* 9: 45-48.
- Murison, David. 1979. 'Norse influence on Scots'. *Lallans* 13: 58-62.

- Murray, James A. H. 1873. *The dialect of the Southern counties of Scotland: Its Pronunciation, Grammar, and Historical Relations*. London: The Philological Society.
- Murray, Mary. 1982. *In my ain words: An East Neuk vocabulary*. Anstruther: Scottish Fisheries Museum.
- National Library of Scotland. 2022. *Maps*. N.p. <https://maps.nls.uk/> [accessed: 2 June 2022].
- Newton, Michael. 2021. *Gaelic in your gob: Four dozen English words that came from the Scottish Highlands*. Ingram: Saorsa Media.
- Nominatim. 2022. *Nominatim API*. N.p. <https://nominatim.org/release-docs/develop/api/Overview/> [accessed: 2 June 2022].
- OED3 = The Oxford English Dictionary*. 2000–. Ed. John A. Simpson. 3rd edn. Oxford: Oxford University Press. <https://oed.com> [accessed: 1 August 2023].
- OpenStreetMap. 2022. N.p. <https://www.openstreetmap.org> [accessed: 2 June 2022].
- Pederson, Lee. 1988. ‘Electronic matrix maps’. *Journal of English Linguistics* 21: 149–174.
- Pödör, Dóra. 1995–1996. ‘The phonology of Scottish Gaelic loanwords in Lowland Scots’. *Scottish Language* 14/15: 174–189.
- Pride, Glen L. 1996. *Dictionary of Scottish building*. Edited by Iseabail Macleod and Pauline Cairns. Edinburgh: The Rutland Press.
- Pröbsting, Lasse. 2020. ‘Galloway lexis: Regional and national influences’. Unpublished BA Dissertation, University of Vienna.
- Riach, William A.D. 1977. ‘A dialect study of comparative areas in Galloway with particular reference to the Irish connection’. Unpublished PhD Dissertation, University of Edinburgh.
- Riach, William A. D. 1988. *A Galloway glossary*. Aberdeen: Association for Scottish Literary Studies.

- Robinson, Chris. 2008. *Say it in Scots: Scottish wildlife*. Edinburgh: Black and White Publishing.
- Robinson, Jonnie. 2021. *A Thesaurus of English dialect and slang: England, Wales and the Channel Islands*. Cambridge: Cambridge University Press.
- Schmitt, L. E. (ed.) 1968. *Germanische Dialektologie: Festschrift für Walter Mitzka zum 80. Geburtstag*, Wiesbaden: Franz Steiner.
- Smith, Jeremy J. 1994. 'Norse in Scotland'. *Scottish Language* 13: 18–33.
- SND = *The Scottish National Dictionary*. 1931–1976. Eds. William Grant and David Murison. Edinburgh: Scottish National Dictionary Association. <https://dsl.ac.uk/> [accessed: 17 April 2023].
- Snodgrass, Catherine. 1943. 'Map of economic regions of Scotland'. *Scottish Geographical Magazine* 59: 15–18.
- Speitel, Hans-Henning. 1969. 'An areal typology of isoglosses: Isoglosses near the Scottish-English border'. *Zeitschrift für Dialektologie und Linguistik* 36: 49–66.
- Speitel, Hans-Henning. 1981. 'The geographical position of the Scots dialect in relation to the Highlands of Scotland'. In Benskin and Samuels (1981): 107–129.
- Speitel, Hans-Henning, and James Y. Mather. 1968. 'Schottische Dialektologie: Zur Terminologie'. In Schmitt (1968): 520–541.
- The Scots Thesaurus*. 1990. Eds. Iseabail Macleod, Pauline Cairns, Caroline Macafee, and Ruth Martin. Aberdeen: Aberdeen University Press.
- Traynor, Michael. 1953. *The English dialect of Donegal*. Dublin: Royal Irish Academy.
- Tulloch, Graham. 1997. 'Lexis'. In Jones (1997): 378–432.
- Upton, Clive, David Parry, and J. D. A. Widdowson. 1994. *Survey of English dialects: The dictionary and grammar*. London: Routledge.
- Withers, Charles W. J. 1984. *Gaelic in Scotland 1698–1981: The geographical history of a language*. Edinburgh: John Donald.

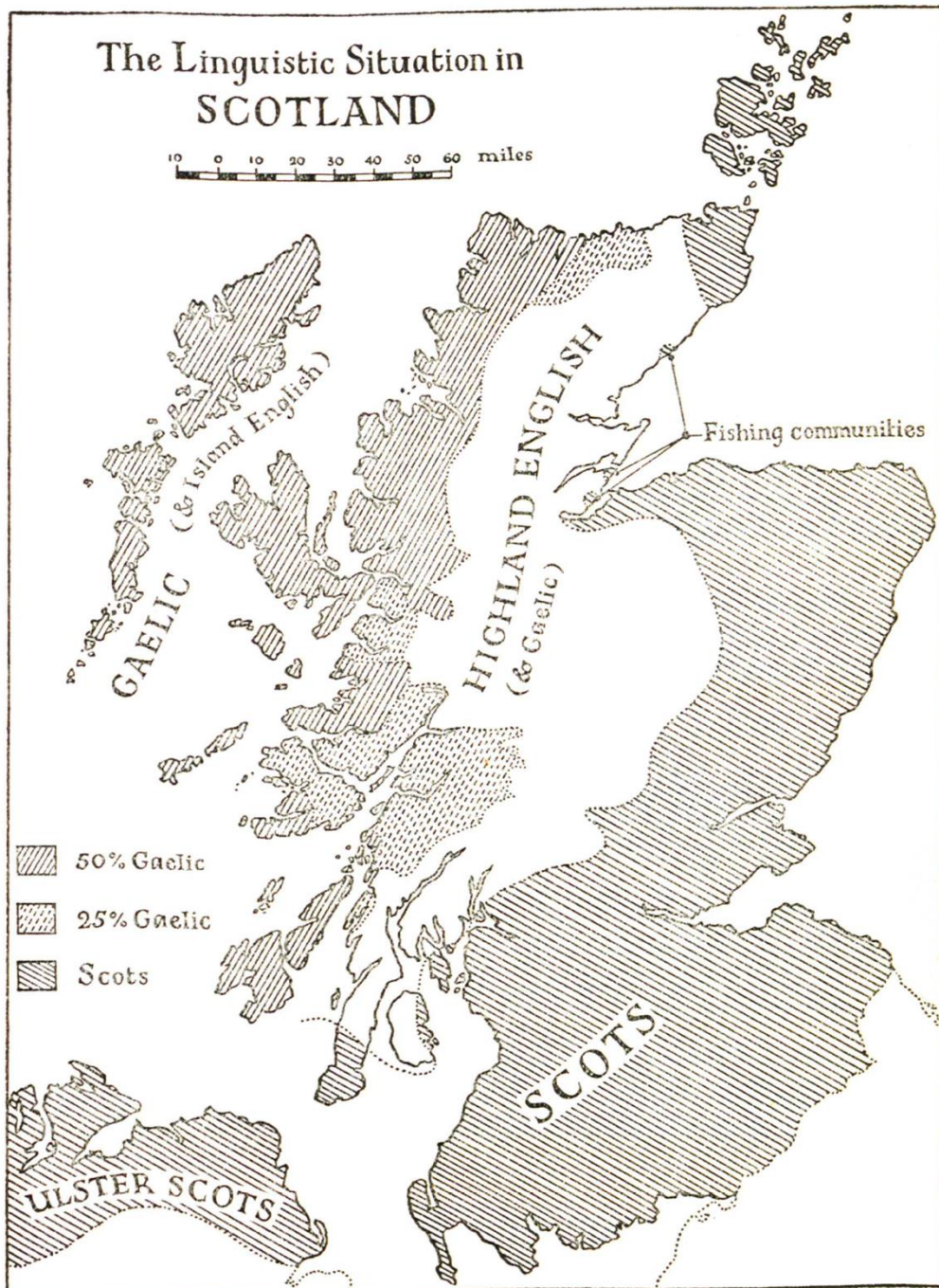
Introduction to the Digital Lexical Atlas of Scotland

Withers, Charles W. J. 1988. *Gaelic Scotland: The transformation of a culture region*. London: Routledge.

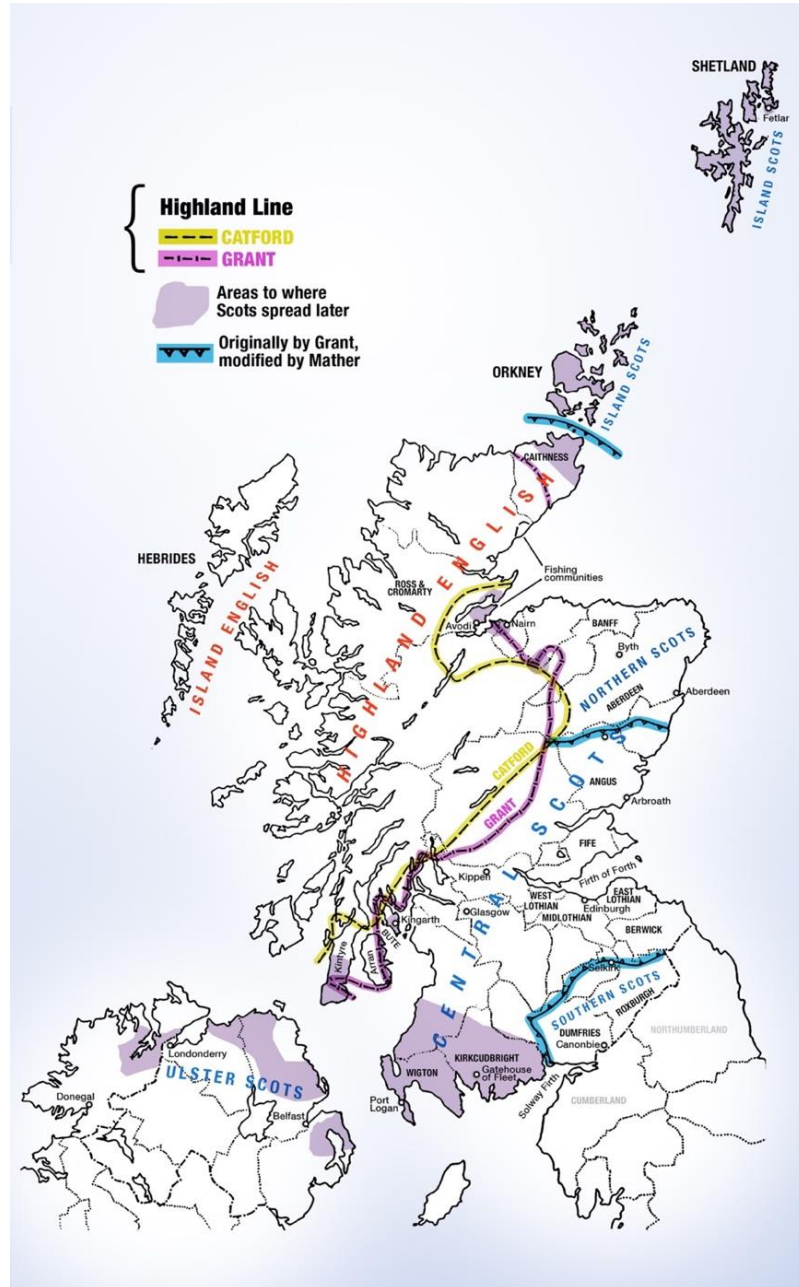
Zwickl, Simone. 1996. *Eine Wortgeographie Nordirlands*. 2 vols. Magisterarbeit. Ruprecht-Karls-Universität Heidelberg.

Appendix

Map 1: The Linguistic Map of Scotland (from Catford 1957) minus Shetland



Map 2: The Linguistic Map of Scotland (redrawn from Speitel and Mather 1968: Karte 1)³⁷



³⁷ Although not shown in Speitel & Mather's map, Grant's line originally includes the shaded areas of and to the North of the Black Isle and runs all the way up the coast to connect with the continuation of the line in Caithness.

Text 1: Excerpt from LAS data, for SPIDER

The extract shows that the original data were organised by lexical variable, and then, within each of these, by county, and then by ortho-phonological variable, each of which is annotated by a numerical locality id and, where relevant, an alphabetic informant number.

84 SPIDER (P01, 178)

Shetland

Creepie craalie - 24

Ettercap - 2, 30

Maamie spider - 17

Paamie - 23

Spider - 1-10, 13-16, 18-20,

21ab, 22, 24-25, 27-33

Nil - 11-12, 26

Orkney

Aiter kep - 15

Ettercap - 11-12, 13b

Kipie kringlie - 13a

Spider - 1-9, 11, 13ab, 15-21

Nil - 10, 14

Caithness

Sheepard - 2a

Spider - 2ab, 5, 8-9, 11, 13-14,

16b

Weaver - 12b

Web spinner - 8

Nil - 1, 3-4, 6-7, 10, 12ac, 15,

16a, 17

Sutherland

Spider - 1, 3-8, 9ab, 10-13, 16

Nil - 2, 14-15, 17

Ross & Cromarty

Pochcan salyin - 31

Sautie pock - 22

Spider - 3-6, 8-9, 13-14, 16-18,

20, 22-24, 25ab, 26-27,

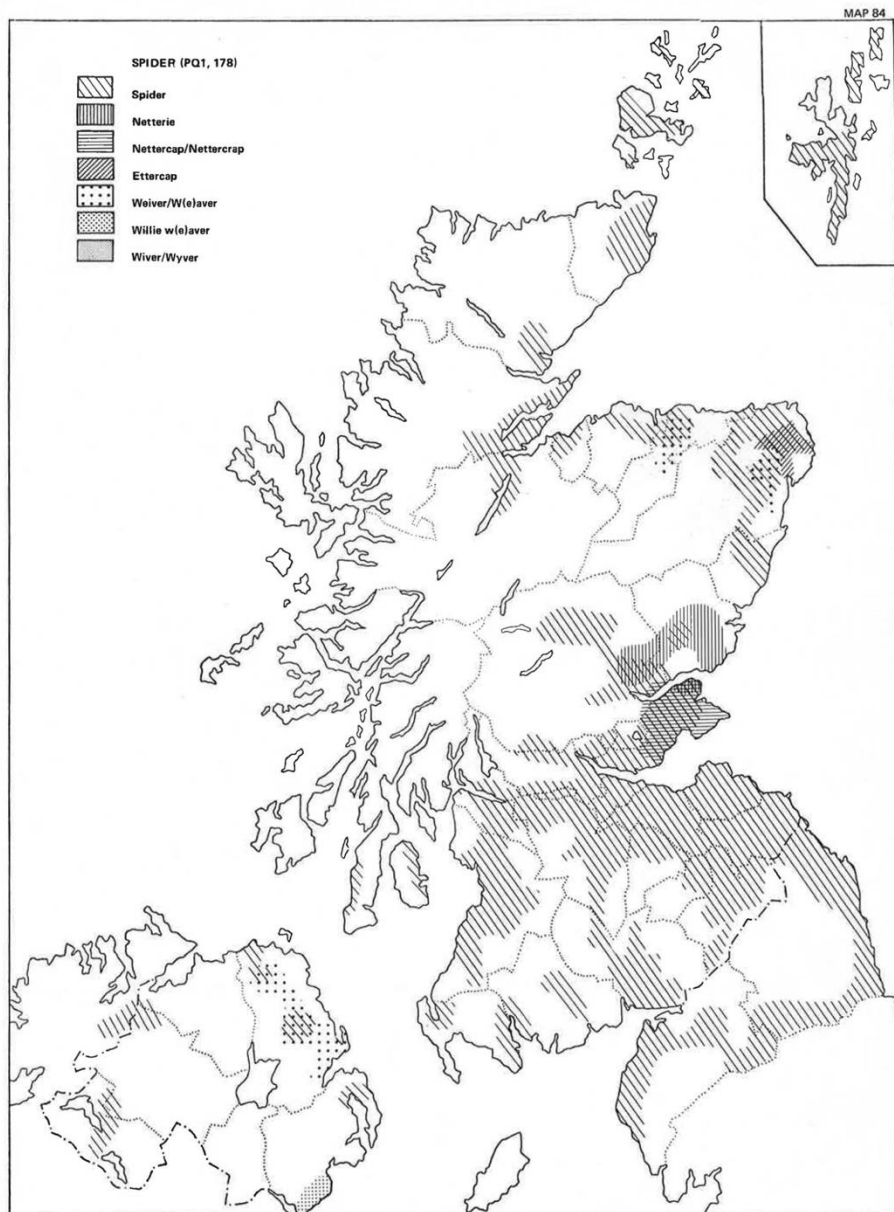
29-31, 32ab, 34, 36, 39

Spinner- 26

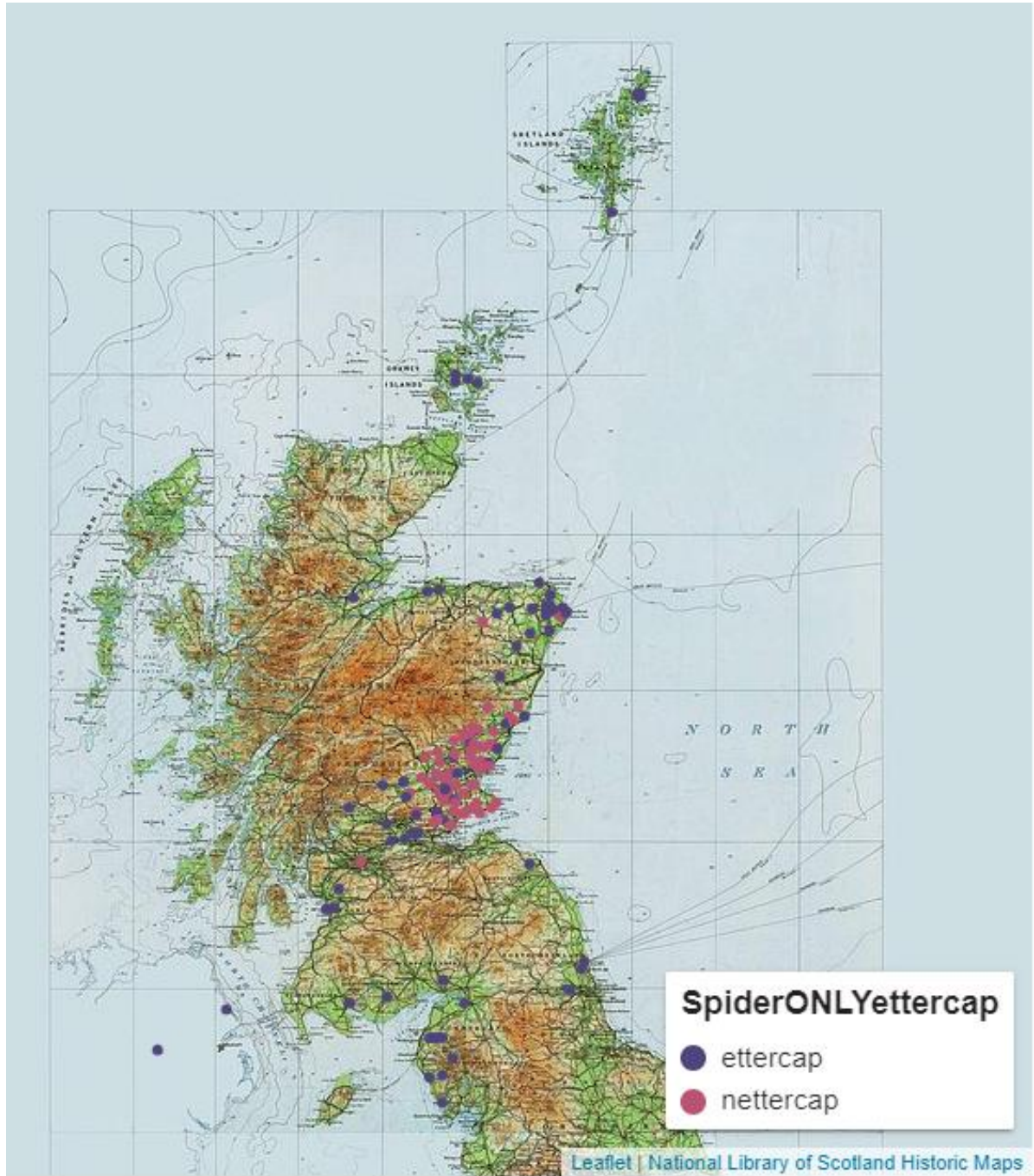
Nil - 1-2, 7, 0-12, 15, 19, 21,

28, 32c, 33, 35, 37ab, 38

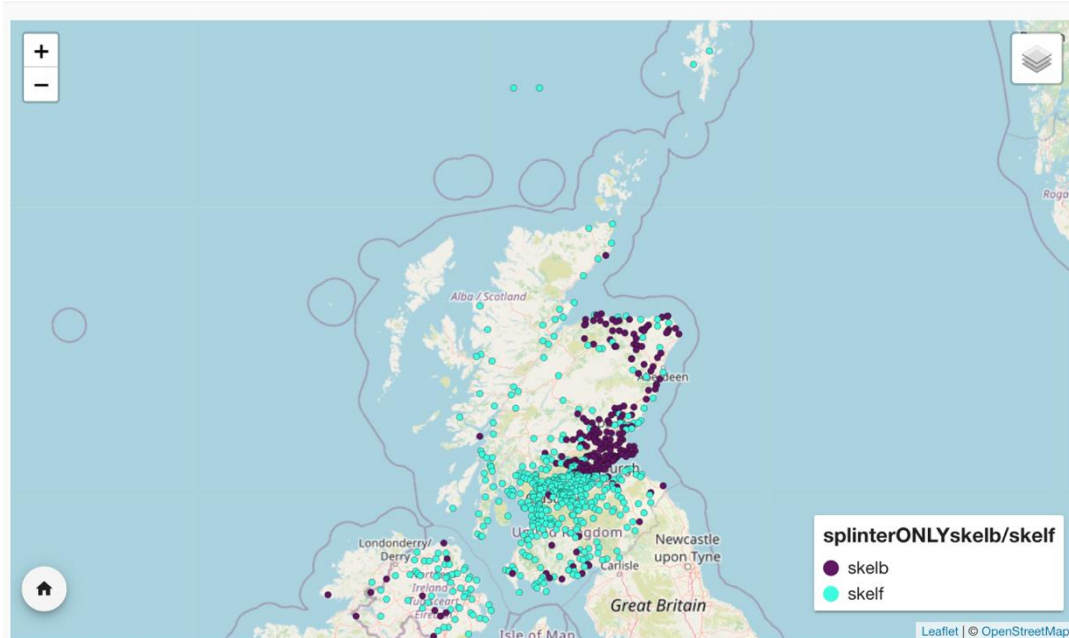
Map 3: LAS map of SPIDER



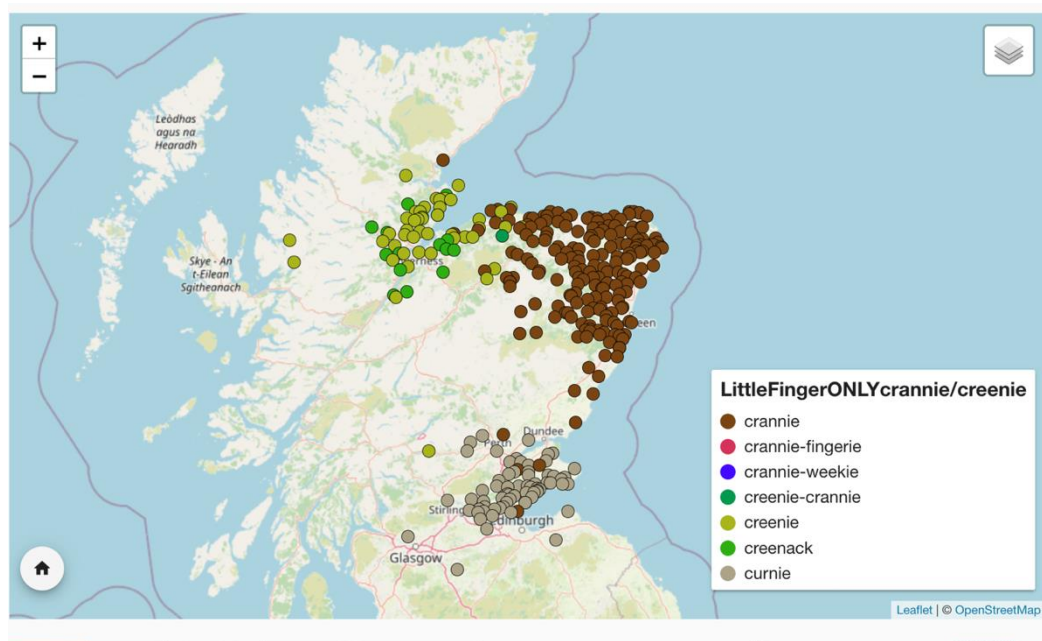
Map 4: *DLAS NLS traditional base ordnance survey map showing the lexeme *ettercap/nettercap* (for SPIDER) by locality*



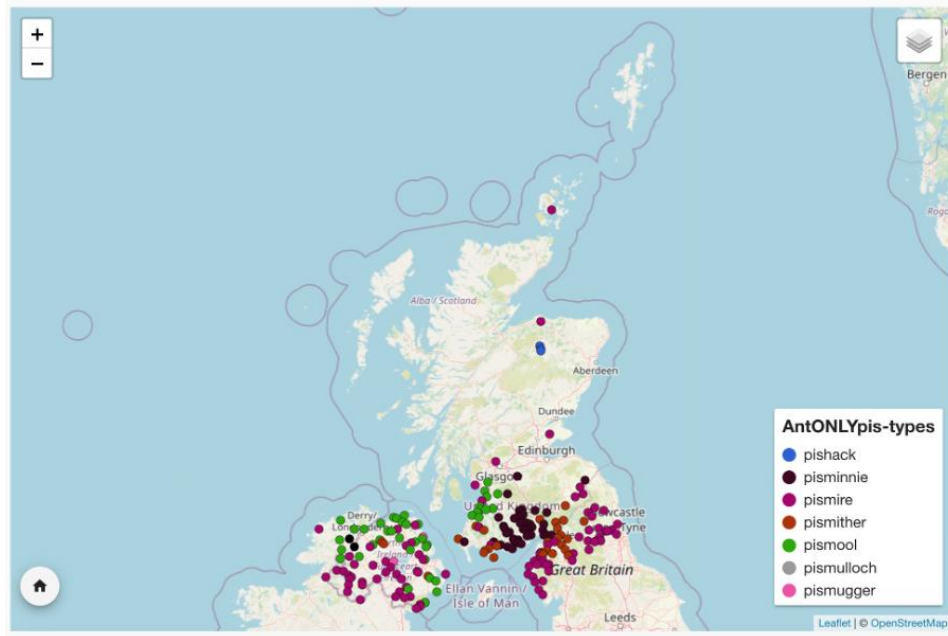
Map 5: *DLAS OpenStreetMap showing all occurrences of **skelb** and **skelf** (for SPLINTER) by locality*



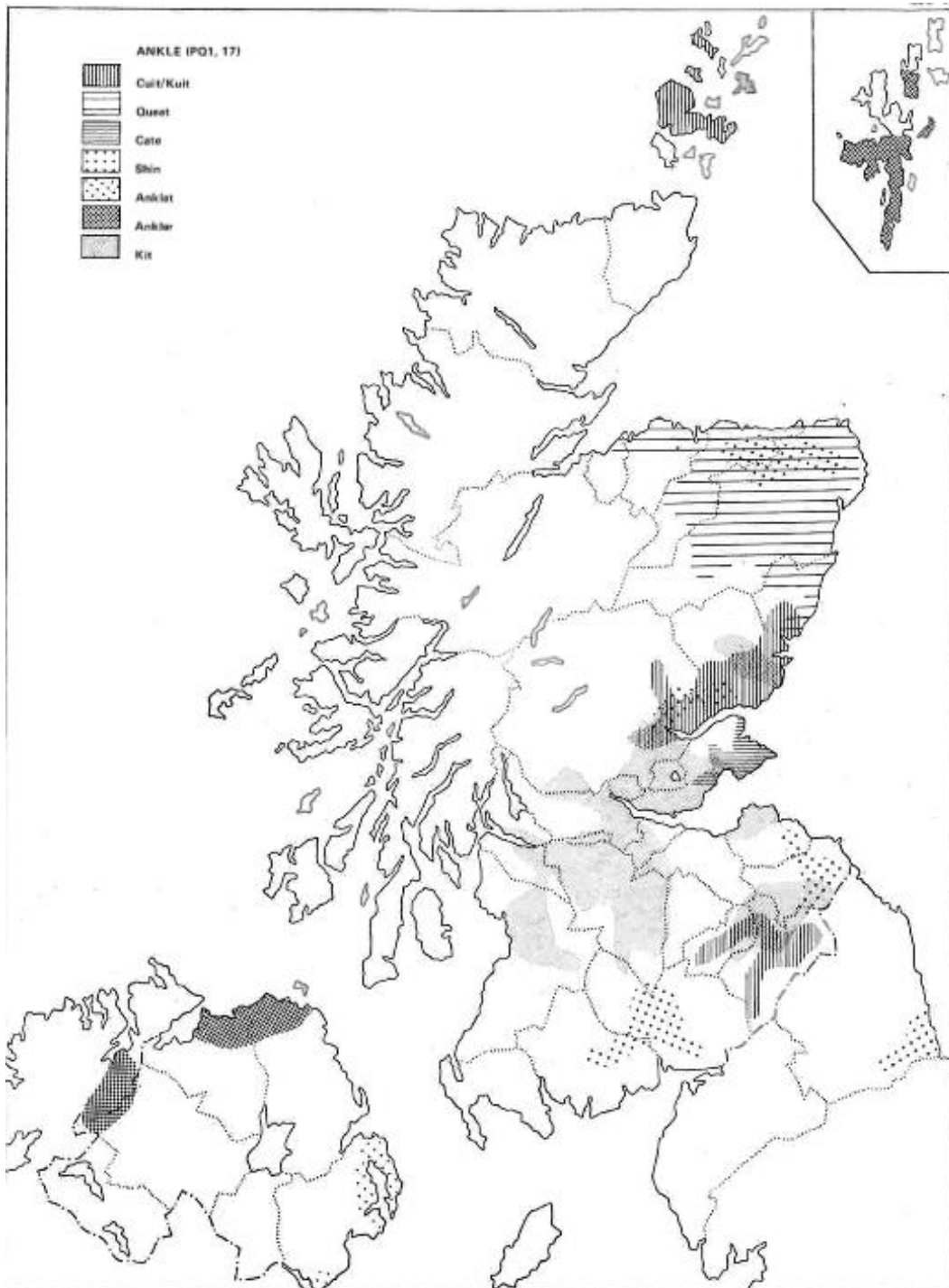
Map 6: *DLAS OpenStreetMap showing all occurrences of **crannie** and **creenie** (for SPLINTER) by locality*



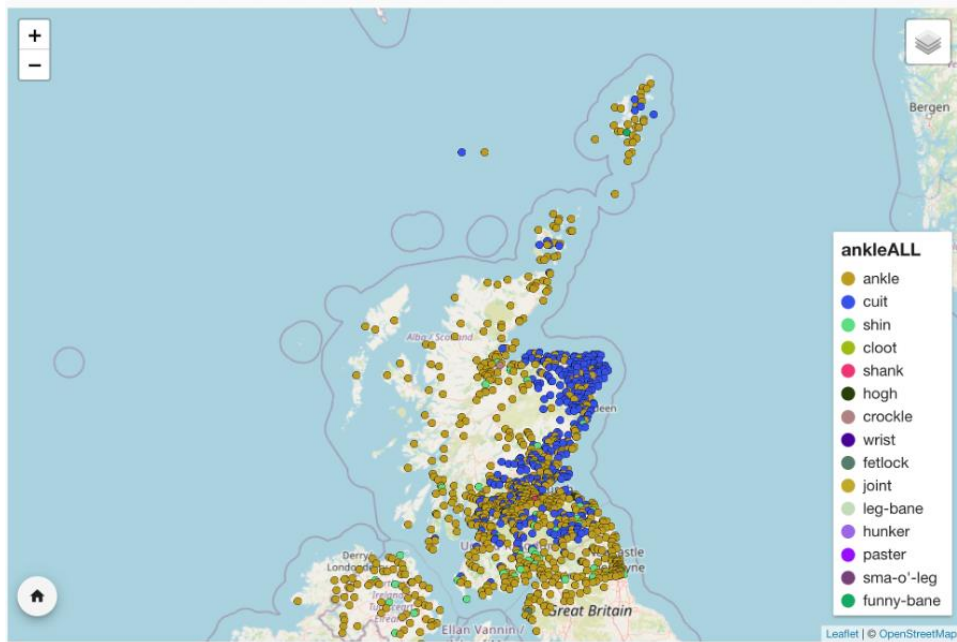
Map 7: DLAS OpenStreetMap showing *pis*-lexeme variants (for ANT) by locality



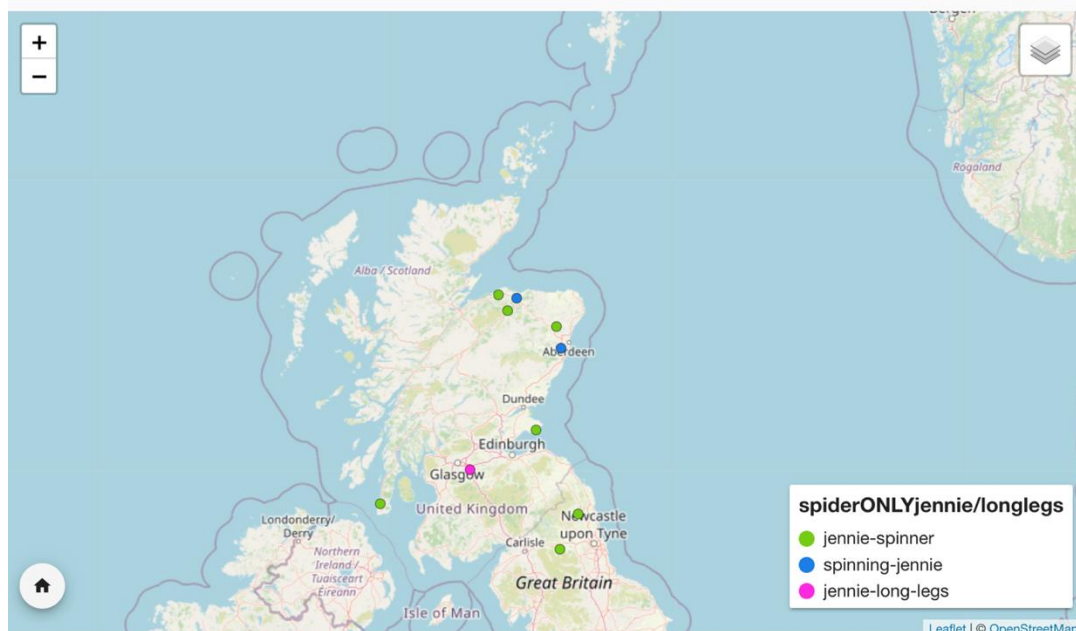
Map 8: LAS Map of ANKLE responses



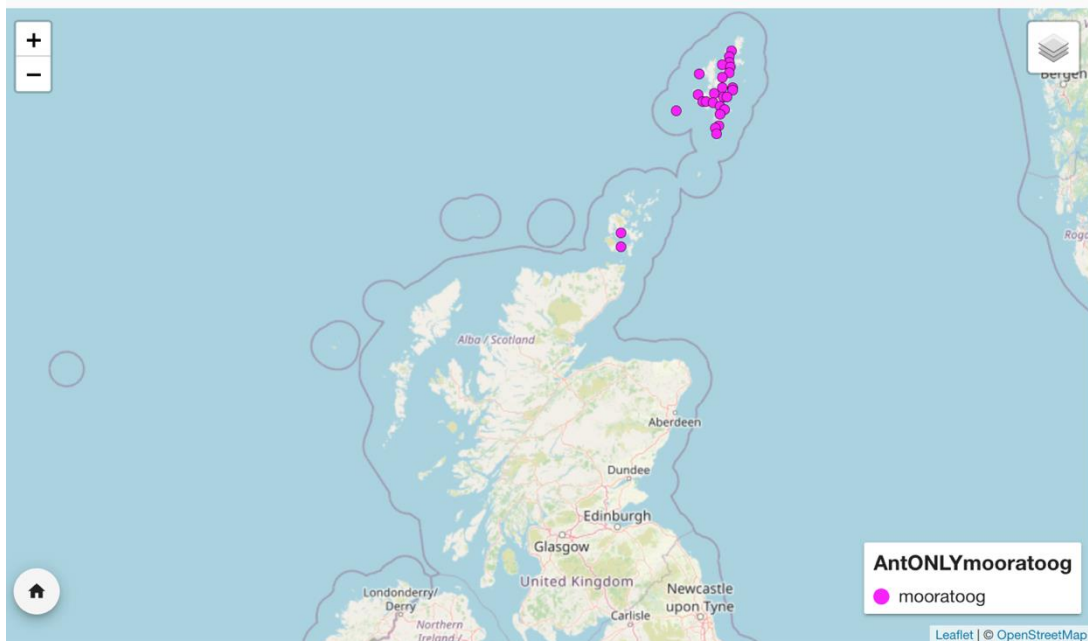
Map 9: *DLAS OpenStreetMap of all lexemes for ANKLE, by locality*



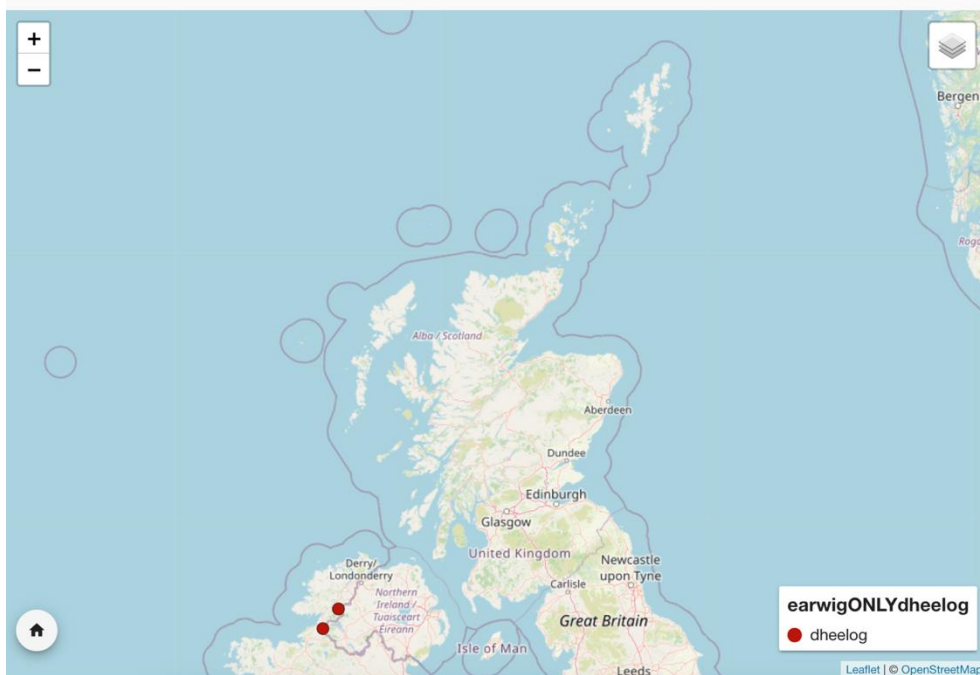
Map 10: *DLAS OpenStreetMap showing the lexeme jennie-compounds (for SPIDER) by locality*



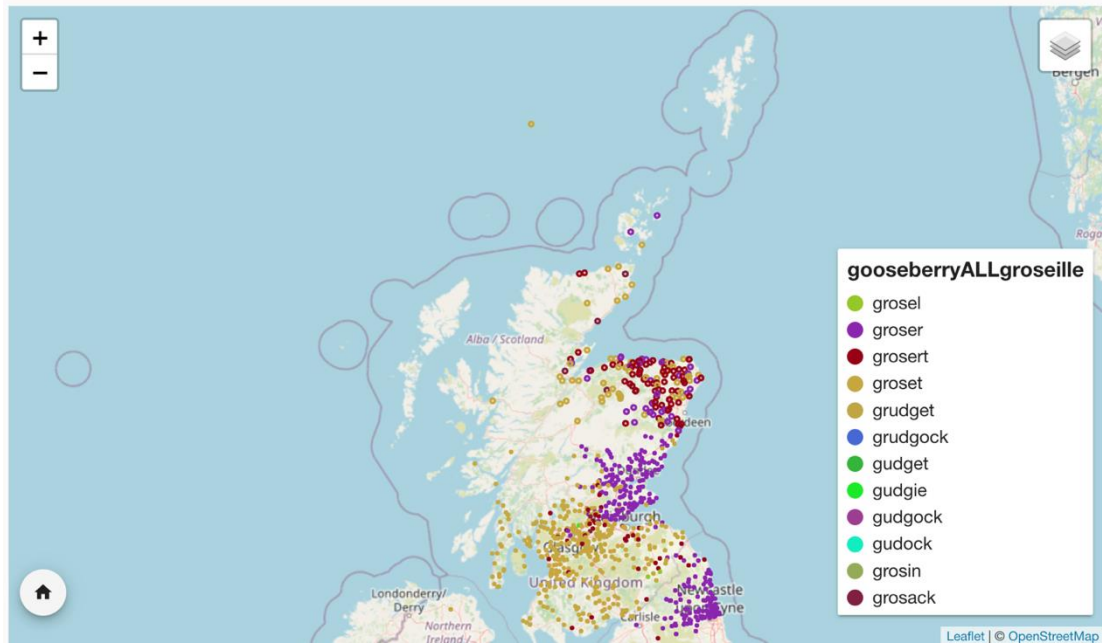
Map 11a: *DLAS* OpenStreetMap showing the lexeme *mooratoog* (for ANT) by locality



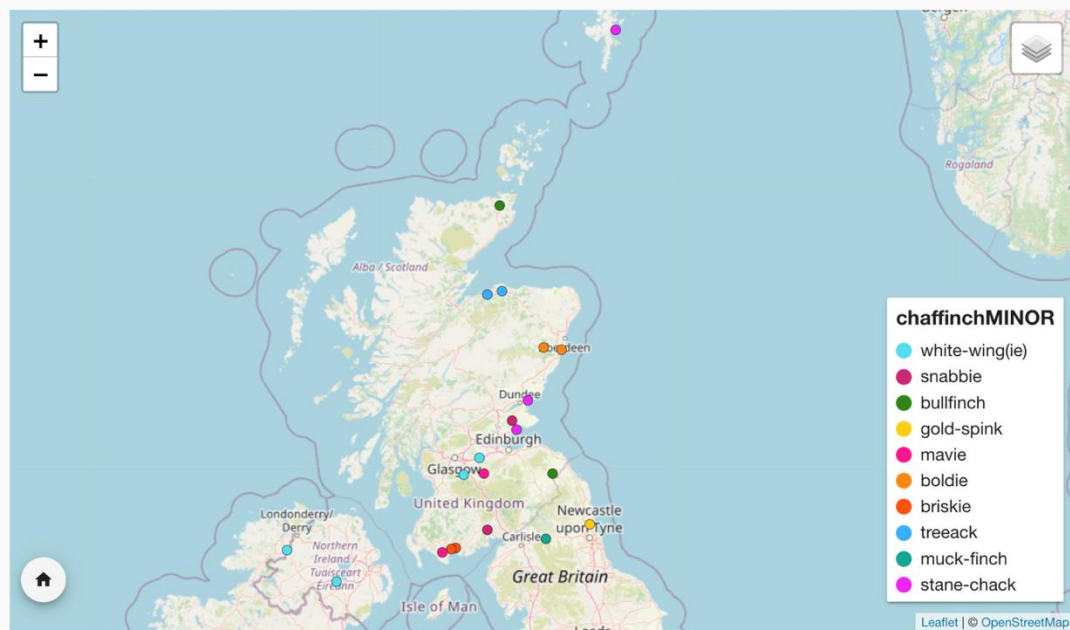
Map 11b: *DLAS* OpenStreetMap showing the lexeme *dheelog* (for EARWIG) by locality



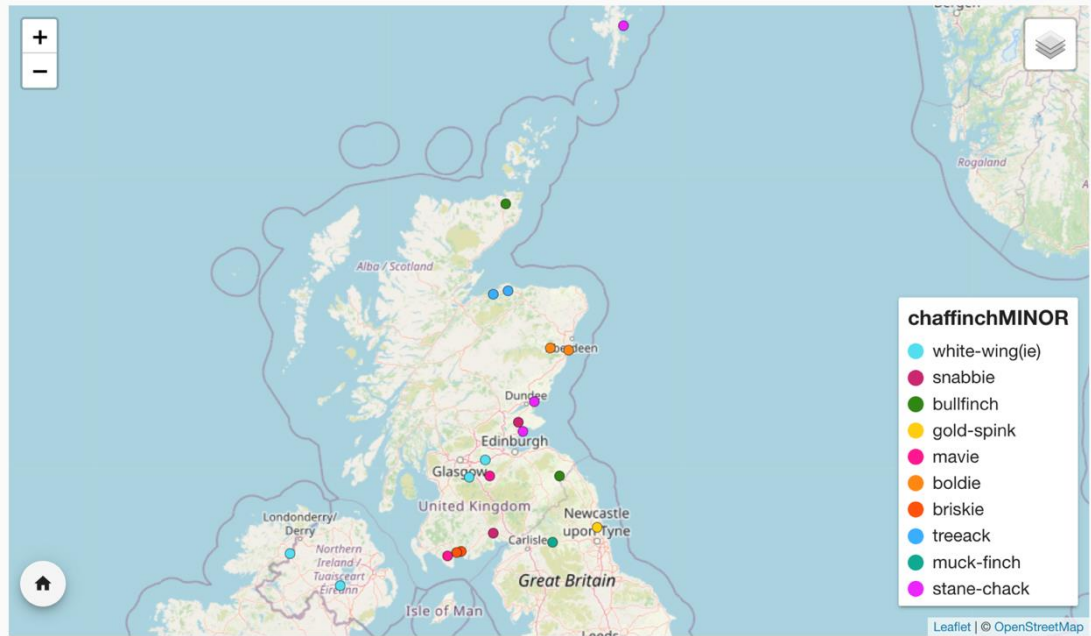
Map 12: DLAS OpenStreetMap showing the all lexemic variants arising from French *groseille* ('gooseberry'), by locality



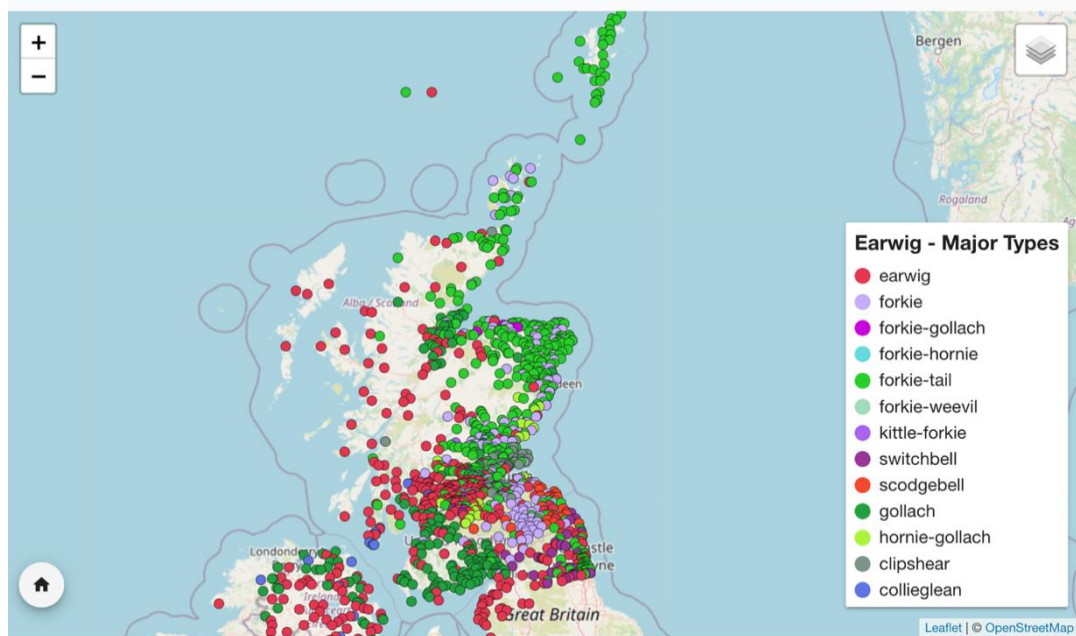
Map 13: DLAS OpenStreetMap showing all oncers for COUCHGRASS, by locality



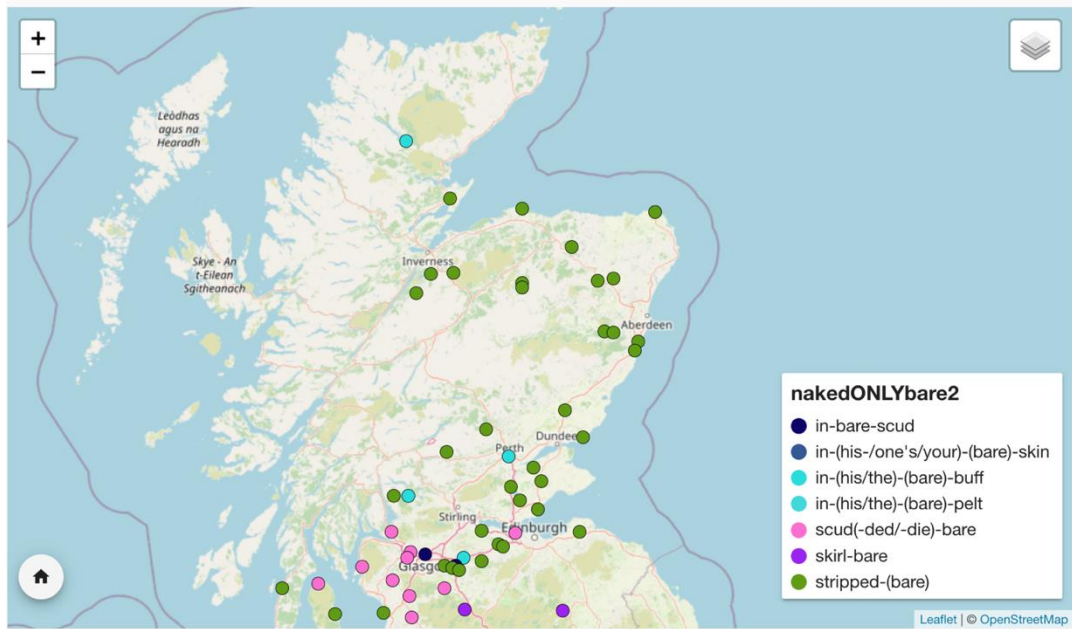
Map 14: DLAS OpenStreetMap showing all minor lexemes for CHAFFINCH, by locality



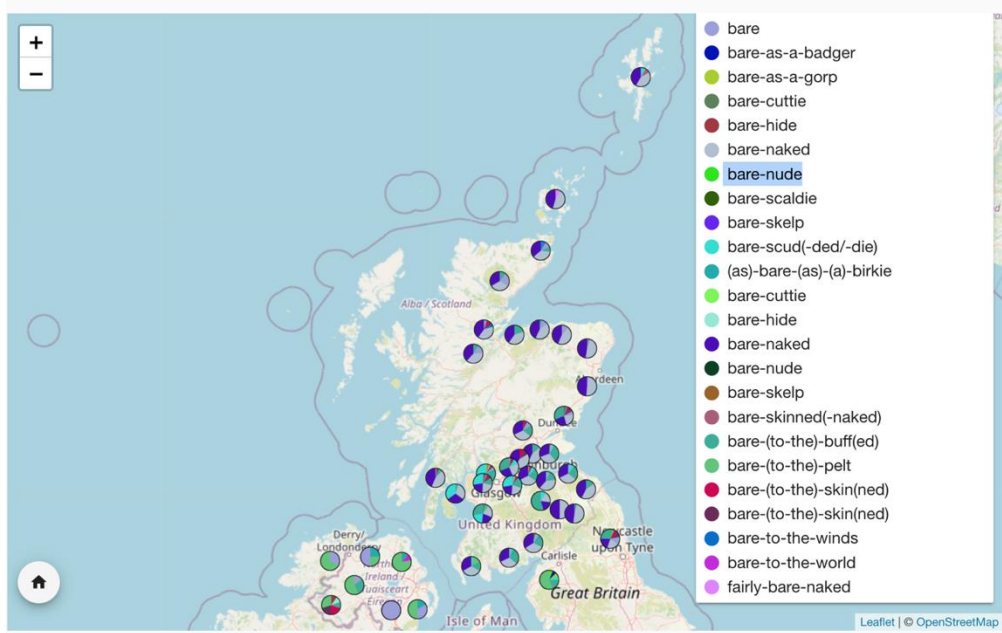
Map 15: DLAS OpenStreetMap showing all major lexemes for EARWIG, by locality



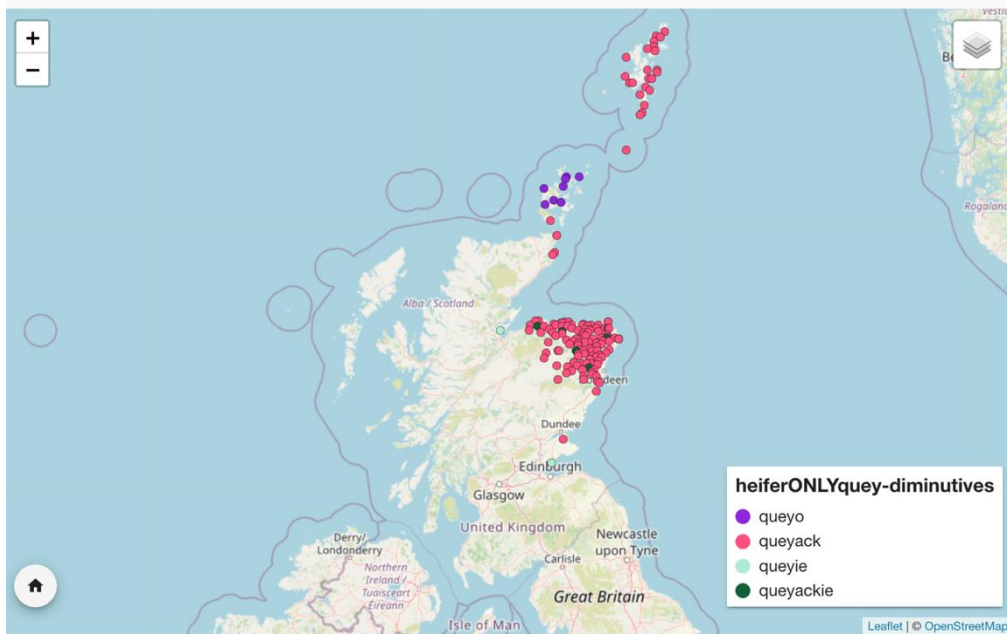
Map 16a: DLAS OpenStreetMap showing a first set of phrases with *bare* for STARK-NAKED by locality



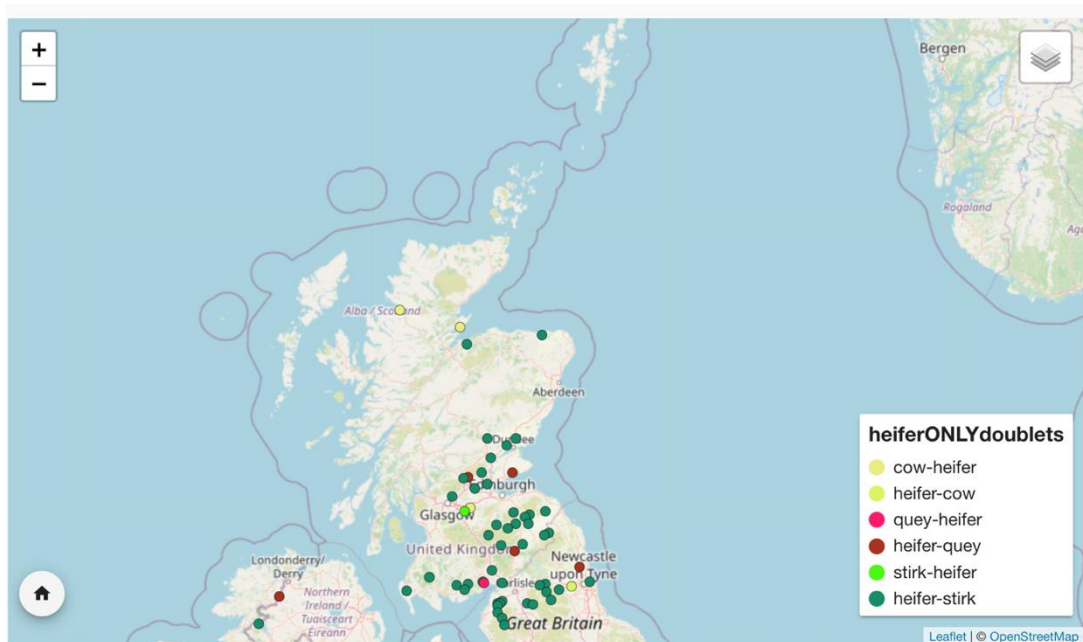
Map 16b: DLAS OpenStreetMap showing a second set of phrases with *bare* for STARK-NAKED by county



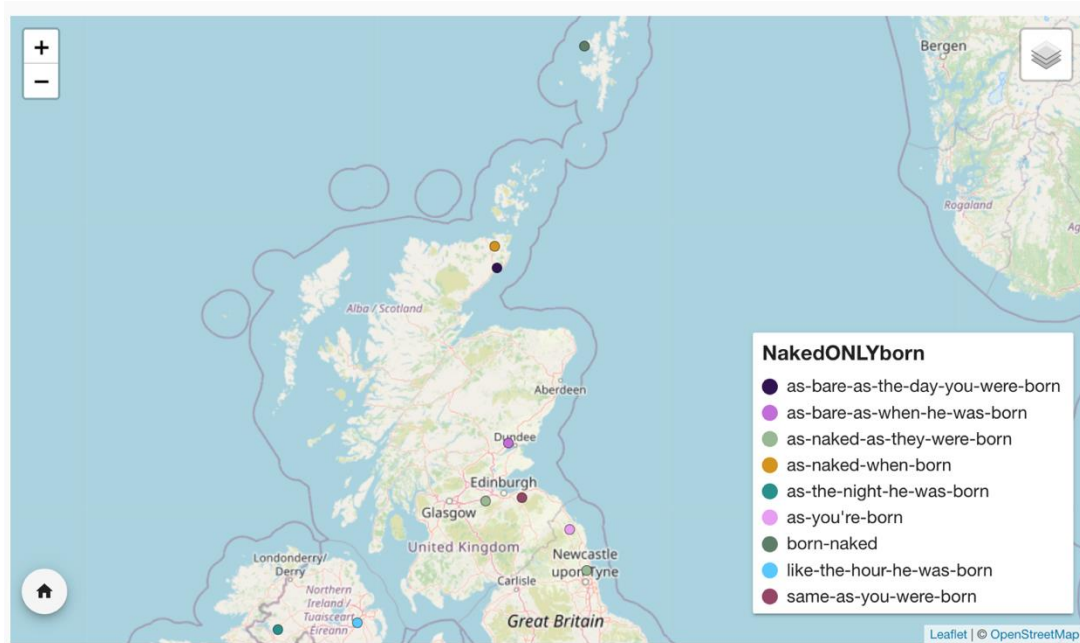
Map 17a: *DLAS* OpenStreetMap showing *quey*-diminutives for HEIFER, by locality



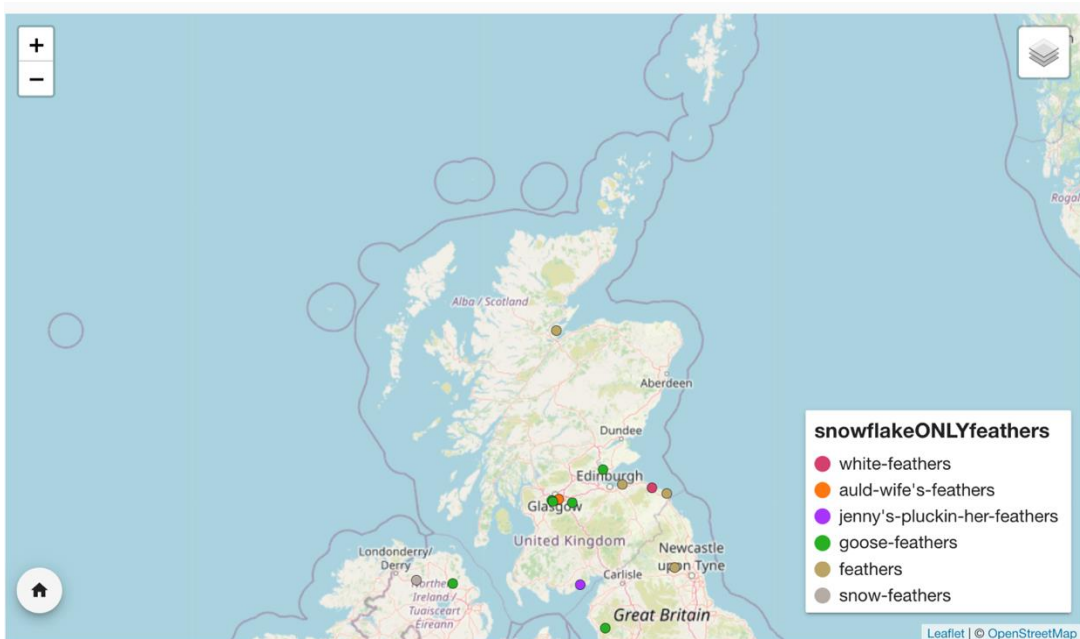
Map 17b: *DLAS* OpenStreetMap showing alternating doublet compounds for HEIFER, by locality



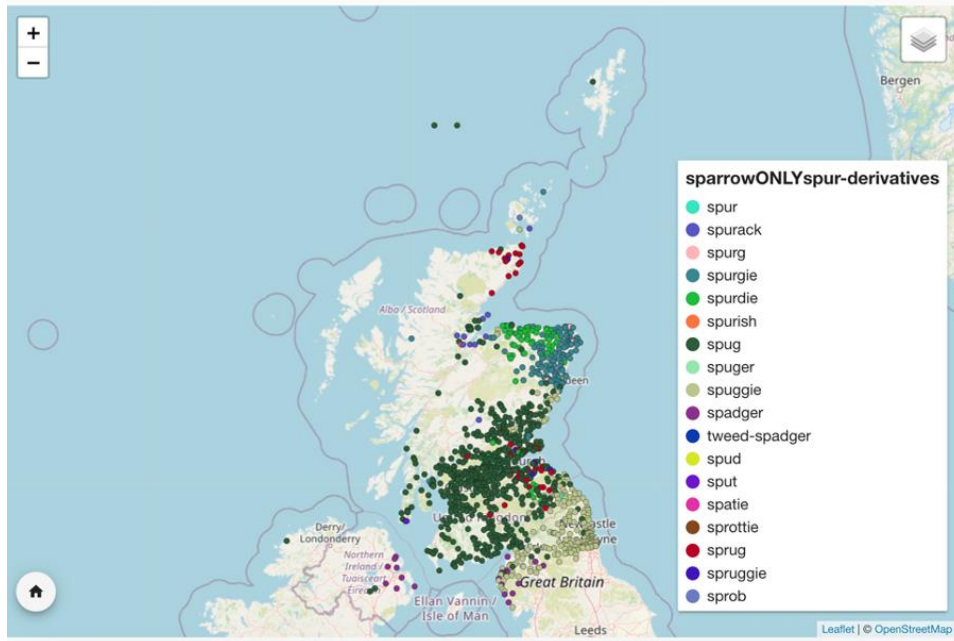
Map 18: DLAS OpenStreetMap showing alternating phrases with *born* for STARK-NAKED, by locality



Map 19: DLAS OpenStreetMap showing words, compounds and phrases with *feathers* for SNOW-FLAKE, by locality



Map 20: DLAS OpenStreetMap Scandinavian-derived forms of *spur* for SPARROW, the ‘passer domesticus’, by locality



Map 21: Economic Map of Scotland, from Snodgrass (1943)



New Economic Map of Scotland.

Map 22: Map of Highland Line by Speitel (1981)

