

RESEARCH

Dialect variation in Scottish Gaelic nominal morphology: A quantitative study

Pavel Iosad¹ and William Lamb²

¹ University of Edinburgh, Linguistics and English Language, Edinburgh, GB

² University of Edinburgh, Celtic and Scottish Studies, Edinburgh, GB

Corresponding author: Pavel Iosad (pavel.iosad@ed.ac.uk)

This paper presents a dialectometric analysis of Scottish Gaelic morphology, with a focus on the noun phrase, using previously unpublished data from the Linguistic Survey of Scotland. Fifty-five morphological features were extracted across 201 survey points, and the data subjected to a variety of analyses, including cluster analysis, regression, and correlation analysis. We establish that the Gaelic noun system shows robust diatopic (i.e. geographical) variation; this challenges previous assertions in the literature, which describe Gaelic morphosyntax as geographically uniform. In addition, we argue that our dialectometric results provide an insight into the grammatical structure of the language that is not easily achievable through the analysis of individual varieties. In particular, we argue that the results highlight a dissociation between morphological categories such as case and gender and their morphophonological exponents, which are particularly complex in Scottish Gaelic. The paper thus serves as a proof of concept for the use of dialectometric findings in a theoretical context.

Keywords: Scottish Gaelic; morphology; dialectology; dialectometry; Celtic studies

1 Introduction

Scottish Gaelic is a Celtic language currently spoken by about 57,600 individuals in Scotland (National Records of Scotland 2015). Although the greatest concentration of speakers today is in the Outer Hebrides, a group of islands off of the west coast, Gaelic was the predominant language over much of the mainland Highlands until quite recently (Withers 1984). Various regional- and single-dialect studies exist, but the Linguistic Survey of Scotland (henceforth “LSS(G)”) is the only broad documentation effort to date. The LSS(G) was conducted in the mid 20th century (Bosch 2006), when speakers still could be found across most of the traditional *Gàidhealtachd* (the Gaelic-speaking area). Despite the availability of dialectal data, the language’s overarching patterns of diatopic variation remain poorly understood. Additionally, available studies are confined mainly to phonetic and phonological variation.

In this paper, we provide a preliminary analysis of unpublished LSS(G) data on nominal morphology, to better understand variation in this understudied domain. We argue that quantitatively investigating the patterns of multiple varieties is not only useful for understanding a language’s dialectal landscape, but can be of significant theoretical import. We make this point by showing how our dialectometric results can help illuminate certain structural aspects of the language’s grammar more readily than even in-depth study of a single variety.

Thus far, dialectometric studies have largely concentrated on phonology and lexis. To an extent, this reflects their underpinning data; older dialect surveys rarely provide

detailed information on other linguistic levels (Wieling & Nerbonne 2015: 256).¹ For those interested in morphosyntactic variation, sparsity issues are compounded by the fact that researchers have tended to regard morphosyntax as less sensitive to geographic patterning (Glaser 2013: 201; Szmrecsanyi 2013: 1). This is evident in the Gaelic context as well: while Hamp (1997: 8) describes the exceptional phonological heterogeneity of Scottish Gaelic dialects, Watson (2010: 118)² and MacInnes (2006: 123–124) emphasise their morphological and syntactic homogeneity, respectively. However, recent work on other languages has shown that diatopic signals at the level of morphology, at least, are similar in magnitude to those of phonology and lexis (Spruit 2008). In some cases, they actually exceed them (Scherrer & Stoeckle 2016: 104).

Three questions concern us here: 1) Does noun morphology in Scottish Gaelic pattern geographically and, if so, how can we explain the patterns of variation? 2) What issues arise when deploying dialectometric methods with the Celtic languages, and Scottish Gaelic specifically? 3) How can a data-rich, aggregative approach help uncover a language's underlying grammatical structure, beyond what is possible from evaluating single features or single varieties in isolation?

In the following sections, we provide preliminary results and interpretation, and confirm the viability of these methods. On the basis of our results, the consensus that Gaelic morphosyntax does not vary diatopically must be revised: this data evinces a clear diatopic signal. This holds even after controlling for variables such as age and gender of the speakers in the questionnaire, and the regional “health” of the language. While the results inform our understanding of the extent of variation within the language, we argue that they also illuminate its underlying grammatical structure. Following Adger (2017), we suggest that examining the covariation of superficially unrelated linguistic features (or the lack of covariation of superficially related ones) can expose some of the structural mechanisms underlying these patterns.

This latter issue is particularly salient for minority language studies. In language contact situations generally, and specifically in the case of Scottish Gaelic, grammatical change is often discussed in terms of attrition and language death (Dorian 1973: 415). Deviations from received notions of “correct grammar” are typically regarded as recent and driven by increasing competence in the dominant language. However, these assumptions are rarely subjected to empirical scrutiny (although see Kennard 2014; 2019; Kennard & Lahiri 2017). To determine their veracity for Scottish Gaelic, we examine correlations between our dialectometric results and census data from the late 19th century, which function as a proxy for the regional health of the language during the formative years of most of the LSS(G) participants. We demonstrate that at least some of the variation observed in Scottish Gaelic morphology shows no obvious connection to patterns of attrition. This has theoretical consequences related to the analysis of the relevant varieties, and also practical ones, in areas such as corpus planning (Bell et al. 2014).

In the remainder of this introduction, we summarise the current state of dialectology and dialectometry involving Scottish Gaelic, and of morphosyntactic dialectometry more generally. In Section 2, we describe the LSS(G) data on which this paper is based and our coding procedures. Section 3 presents the results of several quantitative case studies that speak to the paper’s central aims. The upshot of these results is synthesised in Section 3.6, while Section 4 offers a brief conclusion and some future prospects.

¹ For example, in the *Atlas linguistique de la France*, morphosyntax accounts for only 17% of items (see Goebel 2010: 439).

² Watson (2010: 118) seems to restrict his comment to the dialects of Skye and the Outer Hebrides. However, we find considerable morphological variation even in this “most conservative western” region.

1.1 Gaelic dialectology: Previous work

Studies of Scottish Gaelic dialectology tend to concern either single varieties or isolated features across a range of dialects. No macro-study of Scottish Gaelic dialectology has appeared to date, at any linguistic level. The LSS(G) was the only systematic effort to capture the full extent of the language's geographic variation, and only a fraction of it has been analysed. Barring pioneering papers by Jackson (1967; 1968), most of the inductive work has only occurred since the appearance of the *Survey of the Gaelic Dialects of Scotland*, or SGDS (Ó Dochartaigh 1994–1997), the five-volume collection of contextualising essays and edited phonetic transcriptions from the Survey (see Section 2.1). While certain Gaelic dialects are reasonably well documented, our understanding of the language's macro-variation remains inchoate.

The Scottish Gaelic dialects are at one end of the wider Goidelic continuum, which includes Irish and Manx Gaelic. Only O'Rahilly (1936) has attempted to describe the dialectal divisions of the entire Gaelic-speaking area, but some useful regional studies can be found, such as Borgstrøm's volumes on Skye and Ross-shire (1941) and the Outer Hebrides (1940), and Ó Dochartaigh's study of Ulster Irish (1987). Robertson's (1897; 1898; 1907) early studies of Scottish Gaelic also deserve mention. Most of these studies note morphosyntax in passing, if at all, but some provide observations useful for the present context.

On a phonological basis, Robertson (1907) concludes that only two main dialects, a Northern and Southern one, can be identified conclusively: "It would perhaps be as easy to distinguish thirteen dialects as three".³ Jackson (1968) agrees that the Gaelic dialects form two main divisions, but views their orientation differently. For Jackson, the key division is not north and south, but north-west and south-east, encapsulated in zones which he terms the "centre" (north-west) and the "periphery" (south-east), as shown in Figure 1. Jackson identifies a single central dialect – "on the whole [...] innovating [...] and fairly homogeneous" (Jackson 1968: 67) – while he regards the peripheral dialects as fragmented and more conservative.

Watson (2010) provides an overview of Scottish Gaelic macro-divisions and seemingly contradicts Jackson, by describing the western region as more conservative than the east. Watson is one of the few authors to consider morphosyntactic variation, on which he bases his statement. Several studies have found the peripheral region to be phonologically conservative, but morphosyntactically progressive (e.g. Dorian 1978a; Ó Murchú 1989). Our findings (see Section 3) support this, but, surprisingly, we find high levels of morphological innovation even in certain "central" sub-dialects, such as Assynt and Lewis. This challenges the claim that the entire Hebridean area is morphosyntactically homogeneous (e.g. Borgstrøm 1940: 8; Watson 2010: 108). While we could note additional work on the morphosyntax of single dialects and dialectal sub-regions (cf. Adger & Ramchand 2006; Cole 2015; Dorian 1973; 1978a; b), we are not aware of further scholarship on the macro-variation of Goidelic morphosyntax.

1.2 Morphosyntactic dialectometry: The general context

Morphosyntax has been considered since the beginnings of dialectometry (e.g. Séguay 1971; Goebel 1982), but dedicated studies are relatively scarce (Wieling & Nerbonne 2015: 256). In a programmatic review, Glaser (2013) concludes that we know far less about the dialectal variation of morphosyntax than other linguistic levels. The first dialectometric study of a language's morphology was Heeringa et al. (2009), which examined Flemish and

³ For Robertson, the two most important features were the breaking of historically long *ē (cf. Jackson 1968) and the realisation of historically short vowel + long sonorant combinations (e.g. *dall* "blind").

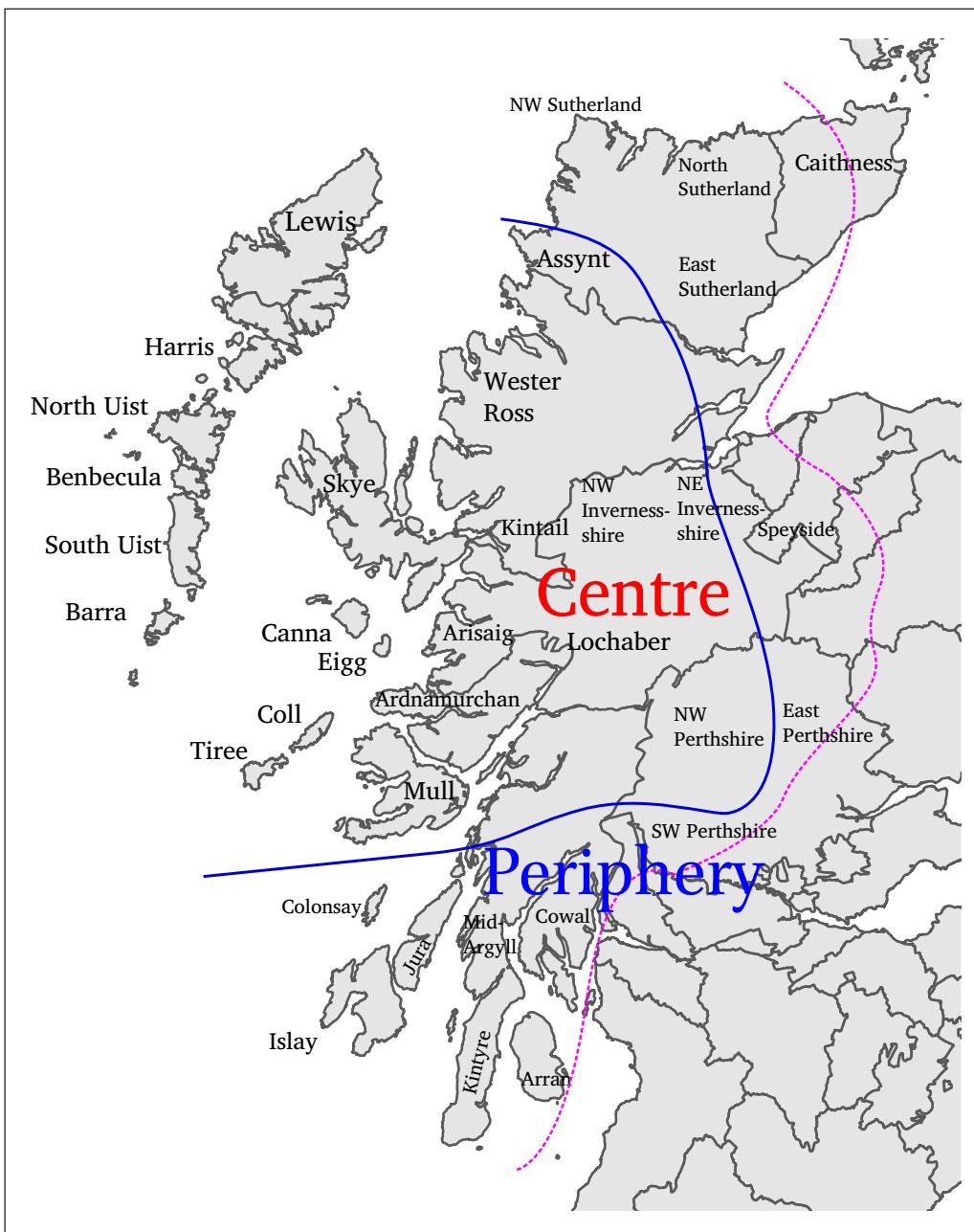


Figure 1: Kenneth Jackson's division of Gaelic dialects. The dotted line shows the approximate eastern boundary of the Gaelic-speaking area in the 1870s (Withers 1984: 92 sqq.).

Netherlandic data from the *Morphological Atlas of Dutch Dialects* (De Schutter et al. 2005; Goeman et al. 2009). Spruit (2008) is the first large-scale investigation of syntax known to us, using data from the *Syntactic Atlas of the Dutch Dialects (SAND)* (van der Ham et al. 2005). He showed that Dutch dialects could be categorised using pairwise measures of syntactic distance and that the results resembled earlier, subjectively-derived atlas maps. He also examined the extent of correlation between different linguistic levels for Dutch dialects, finding that syntax was strongly correlated with phonology, but only weakly with lexis. More recently, Scherrer & Stoeckle (2016) examined the overlap between lexis, phonology, morphology and syntax for Swiss German dialectal variation and measured their individual correlations to geographical distance. In contrast to Spruit (2008), they found that the syntactic data were outliers and less geographically coherent. Morphology was found to have the most geographic coherence (cf. Heeringa & Hinskens 2014).

Szmrecsanyi analysed the geographical variation of English morphosyntax in various publications, using data from the *Freiburg Corpus of English Dialects* (e.g. Szmrecsanyi 2011; 2013; Wolk & Szmrecsanyi 2016). An important finding in Szmrecsanyi (2013: 151) was that, while only 25% of the 57 features examined patterned diatopically on their own, the other 75% contributed to the aggregative analysis: “even geographically seemingly irrelevant features contribute to and are indispensable for an accurate description of the big picture”. Kortmann (2013) used one of the largest datasets of English morphosyntax (*World Atlas of Varieties of English*) to investigate the importance of geographical distance as a predictor of morphosyntactic variation across the sample. Finally, in a study similar to our own, Aurrekoetxea (2016) examined the variation of Basque dialects in terms of nominal morphology.

Perhaps surprisingly, the Goidelic languages have a long, albeit sparse, history of studies that can be described as dialectometric. A pioneer in this respect was Robert Elsie, who conducted dialectometric studies of the lexicon of both the Brythonic (Elsie 1983) and Goidelic (Elsie 1986) languages. His work, however, has not been particularly influential among Celts (Ó Muircheartaigh 2014). On the other hand, Kessler (1995), who worked on Irish Gaelic, is notable for pioneering edit distance in dialectometry; his study has been influential across the discipline at large.

2 Data and methodology

The data for the present study derives from the unpublished “morphophonological” part of the LSS(G) questionnaires. In the first subsection below, we provide a brief overview of the Survey, including its form, demography, limitations and motivational background. Following this, we discuss the nature of our data, with a brief overview of Gaelic nominal morphology. Finally, we describe our coding procedures and analytical methods.

2.1 The LSS(G) dataset

The Linguistic Survey of Scotland was undertaken at the University of Edinburgh from 1949, with separate sections for Scots and Gaelic. The history of the Gaelic section is discussed in detail by Gillies (1997); cf. Jackson (1958) and Bosch (2006). The original materials gathered for the LSS(G) are housed in the School of Scottish Studies Archives. They include questionnaire returns, transcriptions of continuous speech, tape and digital audio, fieldwork diaries, administrative records, palatograms, incomplete questionnaires from the most peripheral regions (such as Nairn and the Trossachs), and other ancillary materials.⁴ However, the principal published output of the project to date has been the phonetic material edited by Ó Dochartaigh (1994–1997).

The architect of the LSS(G) was Professor Kenneth Hurlstone Jackson (1909–1991) and the questionnaire’s orientation reflects his primary interest in historical phonology. Morphology is surveyed in part, but lexis and syntax are largely ignored. The informant selection process balanced sometimes competing concerns for fluency and geographical spread (Gillies 1997: 34). Marginalia in the questionnaires, such as “not spoken Gaelic for 40 years”, indicate that compromises were occasionally made to expand coverage. Consequently, some of the participants were not fluent enough for our purposes (see Section 2.2).

The LSS(G) was biased towards the demographic that some recent literature identifies as NORMs – non-mobile, older, rural males (Chambers & Trudgill 1998: 29). For instance, of the 201 informants included in this study, 130 (65%) were male and 71 (35%) were female. Further information on demographics and inclusion criteria is provided in Appendix A in

⁴ These materials have recently been catalogued, thanks to a grant from the John Lorne Campbell Legacy by way of *Faclair na Gàidhlig*.

the Supplementary Material. Having now provided a contextual overview of the Survey, we shall describe the morphophonological section in more detail and briefly outline the aspects of Gaelic grammar that concern us here.

2.2 Relevant aspects of morphology

The primary data for the present study came from the section “Nouns and Adjectives”, found on pages 38–39 of the LSS(G) morphological questionnaire. In this section we briefly describe the relevant aspects of Gaelic nominal morphology; for additional clarification, see Lamb (2008: 202–213); Gillies (1993: 254–262); Cox (2017).

Gaelic nouns and adjectives decline for case, definiteness, gender and number. Distinctions are maintained through a variety of morphological and morphophonological processes:

- Suffixation of segmental material: *beag* [pɛk] ‘small.NOM.SG’ ⇒ *beaga* [pɛkə] ‘small.NOM.PL’;
- Palatalisation (or depalatalisation) of final consonants: *fraoch* [fruːx] ‘heather.NOM.SG’ ⇒ *fraoich* [fruːç] ‘heather.GEN.SG’;
- Change in the quality of the stem vowel, which may (but does not necessarily) accompany the palatalisation of the final consonant:
 - Short vowels: *cnoc* [kʰrōxk] ‘hill.NOM.SG’ ⇒ *cnuic* [kʰrūçk] ‘hill.GEN.SG’;
 - Long vowels: *ceòl* [kʰjɔ:l] ‘music.NOM.SG’ ⇒ *ciùil* [kʰju:l] ‘music.GEN.SG’
- Initial consonant mutation, which is a complex set of changes that affect the initial consonants of words, usually triggered by the presence of a particular lexical item before the word: *bean* ['pɛn] ‘(a) woman’ ⇒ *a' bhean* [ə 'vɛn] ‘the woman’. There are several kinds of mutation:

Lenition involves various featural changes, in particular from a stop to a fricative, as in the case of NOM.SG *an cat* /ən kʰatʰ/ “the cat” ⇒ DAT.SG *leis a' chat* /leʃ ə xatʰ/ “with the cat”. Orthographically, lenition is represented by adding < h > after the affected consonant.

“**t-sandhi**” is another type of initial mutation. It applies after the definite article to vowel-initial masculine nouns in the nominative (e.g. *àl* /a:l/ ‘brood’ ⇒ *an t-àl* /ən tʰa:l/ “the brood”), as well as to s-initial nouns in various contexts, in which case the /tʰ/ replaces the /s/: cf. *sùil* /su:l/ ‘eye’ ⇒ *leis an t-sùil* /leʃ ən tʰu:l/ “with the eye”).

Nasalisation is a sandhi process that applies to word-initial consonants, particularly stops, following nasal-final proclitics, such as the definite article *an*. Significant variation obtains in how stops are realised in this context, described in detail by Bosch & Scobbie (2009) with SGDS material. Here, we have only included the nasalisation of /s/ (*sùil* /su:l/ ‘eye’ ⇒ *an sùil* /ə zu:l/ “the eye”). This is a less common variety of nasalisation, which is associated with innovative grammar and found only in mainland areas.

- Suppletion: a small number of nouns show highly irregular paradigms that do not fall within the commonly recognised morphological patterns. In our data, this applies to the nouns *bean* ‘woman’ and *bò* ‘cow’.

These processes combine fairly freely with one another within a paradigm, so that forms of a single word may bear little superficial resemblance to each other: cf. *cas* [kʰas] ‘(a) foot.NOM.SG’ with *chois* [χɔ:s] ‘foot.DAT.SG.DEF’. Traditionally, the combined effects of stem-final consonant palatalisation and concomitant vowel raising or fronting are referred to as *slenderisation*. In this study, we focus on the morphological categories of case, gender, definiteness, number and declension class (for additional clarification, see Appendix B).

For the sake of concreteness, we adopt a piece-based, Item-and-Arrangement view of morphological structure, coupled with a Generalized Nonlinear Affixation (Bermúdez-Otero 2012) approach to morphophonology. We assume, in particular, that the phenomena described in this section mostly derive from the concatenation of the root with segmental or subsegmental material. This is reasonably straightforward in the case of suffixation, but is more controversial for the apparently nonconcatenative processes of slenderisation and initial mutation; see notably Stewart (2004) for extended discussion of the challenges that the Gaelic mutation systems poses for morpheme-based theories, and Hannahs (2011) for an overview of the Celtic mutation processes from the perspective of phonological theory. Nevertheless, we will assume that both mutation and slenderisation could be analysed by means of the affixation of some floating subsegmental material to either the left or the right edge of the item undergoing the process.

At least in the case of the short vowels, some authors have argued that the change in the vowel can be an automatic phonological consequence of the palatalisation of the following consonant (e.g. Ó Maolalaigh 1997) rather than a morphological operation. If this analysis is sustainable, it further supports the idea that the manifold effects of “slenderisation” are to a large extent reducible to the single operation of concatenation with a palatalizing phonological feature.⁵ In the case of initial mutations, a Generalized Nonlinear Affixation analysis for Gaelic would require the postulation of floating manner features. Although we are not aware of work specifically offering such an analysis for Scottish Gaelic, the general approach previously has been applied to mutation systems of the other Celtic languages (e.g. Swingle 1993; Iosad 2014; Breit 2019), and we would expect it to be transferable to Scottish Gaelic without much difficulty.

Under this régime, the nonconcatenative morphophonological phenomena described in this section can be analysed as involving lexical entries of a similar kind to those deployed for affixation. Consider, for instance, the formation of adjective plurals. The basic rule is that monosyllabic adjectives form their plural by the addition of [ə] (without any slenderisation or depalatalisation), whilst for adjectives of more than one syllable, the plural form is identical to the nominative singular.⁶ This could be represented as follows:

- (1) Lexical entry for adjective plurals

$$[\text{PL}]: \left\{ \begin{array}{l} \sigma\sigma_- \Leftrightarrow \emptyset \\ \Leftrightarrow \text{ə} \end{array} \right\}$$

The lexical entry shows the morphosyntactic category (PLURAL), and the subcategorisation frames for the possible exponents: zero if the exponent is preceded by (at least) two syllables, and [ə] in all other cases. In the exact same way, the exponence of the dative case in the singular depends on the gender of the noun: feminine nouns undergo slenderisation, but masculine ones do not. The lexical entry is then as in (2).

- (2) Lexical entry for dative singulars

$$[\text{DAT SG}]: \left\{ \begin{array}{l} [\text{FEM}] \Leftrightarrow ^j \\ \Leftrightarrow \emptyset \end{array} \right\}$$

⁵ This approach has long been widely accepted for the closely related Irish; see Ó Maolalaigh (1997); Anderson (2016) for extensive discussion. However, it is equally clear that not *all* effect of slenderisation can be so analysed: at the very least, a different account is probably needed for cases like *cíuil* ‘music.GEN.SG’, where it affects long vowels.

⁶ For simplicity’s sake, we abstract away from further complications involving initial mutation here.

The formalism we adopt here permits a clear separation between the morphosyntactic category being expressed, the contextual factors that influence its exponence, and the exponent itself – a distinction that will be important in our theoretical discussion in Section 3.5.

2.3 Questionnaire data

Turning to the data itself, Table 5 in Appendix D provides a list of the prompts, with glosses and alternate lexemes. Four slightly different versions of the questionnaire were used in the Survey. As there had been no pilot stage, the questionnaire evolved in response to fieldwork and the project's goals (Ó Dochartaigh 1994: vol. 1, p. 57). Apropos of the morphophonological section, versions varied slightly regarding which lexical items were used for some morphological classes, but the overall structure was static. We excluded from the analysis any candidate feature which could have been affected by lexical discrepancies. For example, some Type V feminine nouns began with consonants (e.g. *cathair* 'chair') and some began with vowels (e.g. *uair* 'hour'); therefore, we did not have representative data to investigate lenition after the definite article in the nominative (*a' cathair* vs *an cathair*) for this declension, since lenition only applies to initial consonants.

2.4 Coding procedures and feature selection

The questionnaire responses were entered into a spreadsheet organised by respondent (rows) and prompt (columns). Every feature of interest apart from nasalisation,⁷ can be represented by normal Gaelic orthography. Consequently, during data entry, responses (a total of 11,856 cells) were transliterated from IPA to standard spelling. This was useful information loss given the morphological purpose of the study, and it facilitated automatic feature extraction (see below). In a small number of cases, we normalised a datum towards a single lexical item to enable comparability, but the morphological information was always preserved in the normalised form. For example, due in part to the different versions of the questionnaire, respondents used six different Type V feminine nouns (*acair* 'anchor', *cathair* 'chair', *caora* 'sheep', *iuchair* 'key', *nathair* 'snake' and *uair* 'hour'), which were modified by three different adjectives (*beag* 'small', *crion* 'small' and *mòr* 'large'). In this case, we normalised all nouns to CATHAIR, and all adjectives to BEAG. So, for instance, the return *na h-uarach mòire* 'of the large hour' was normalised as *na cathrach bige* 'of the small chair'.

Following this procedure, we extracted as many features as possible from each response using IF statements in Excel. Taking the above example (Type V feminine nouns in the genitive singular), we extracted four features, i.e. *naⁱ cathrachⁱⁱ biⁱⁱⁱge^{iv}*, as follows:

- i. Is the definite article in the feminine genitive singular form *na*?
- ii. Does the noun show the genitive form *cathrach*?
- iii. Does the adjective show slenderisation, i.e. raising of the stem vowel and palatalisation of the final consonant (*bige*)?
- iv. Is the suffix [ə] present on the adjective (*bige*)?

The data were categorical, with three coding possibilities: feature present ("1"), feature not present ("0") or null (blank cell). If a respondent gave two forms, one of which evidenced the feature, we took an optimistic view and considered it present. For instance, if a respondent provided both *na cathrach bige* (slenderised) and *na cathrach beaga* (non-slenderised) in the genitive, we coded the adjective as slenderised.

⁷ To represent nasalisation, a superscript 'n' was inserted before the nasalised consonant (e.g. *nan* "cat beaga 'of the small cats').

In total, we examined 55 features across 201 respondents. The full list of features is shown in Table 4 in Appendix C. We refer to them using a notation that encloses the morphophonological element of the prompt concerned by the feature in [square brackets]: for instance, for the prompt *naⁱ cathrachⁱⁱ biⁱⁱⁱge^{iv}* as described above, the shape of the article (feature (i)) is notated as [na] cathrach bige, while the presence of slenderisation in the adjective (feature (iii)) is notated as na cathrach b[i]ge. The full list of prompts, showing the features exemplified by each one, is described in Table 5 in Appendix D.

3 Results

In this section we report the results of two strands of work on the data described in Section 2. First, we analyse *diatopic* variation, that is, differences in morphological patterning among the survey points, in order to verify the existing knowledge on Gaelic dialect divisions. This is a core dialectometrical task, to which numerous sophisticated methods have been applied in the literature (e.g. Grieve 2014; Wieling & Nerbonne 2015; Scherrer & Stoeckle 2016). To visualise overall patterns of diatopic variation, we use agglomerative clustering based on the processed data (specifically, edit distance), and hierarchical clustering on principal components extracted from the morphological features. The results indicate that our data is, in principle, suitable for dialectometric enquiry, and adduce confidence for the more theoretically inclined propositions that follow.

We then discuss differences in the patterning of morphological features themselves. First, we examine how geographical patterning of features interacts with demography, analysing it using nonlinear regression (specifically generalized additive modelling) and show how this interplay potentially allows us to disentangle “endogenous” variation from changes driven by attrition. Second, we use correlation analysis to discover clusters of features that behave similarly across varieties of Scottish Gaelic, and argue that many such clusters can be interpreted as demonstrating the workings of the underlying grammar (Adger 2017).

3.1 Aggregate measure of conservatism

Jackson’s choice of features for the morphological sections of the LSS(G) was wholly intentional; he selected those belonging to a perceived grammatical ideal known sometimes as “Biblical Gaelic”. Most of the norms of this variety stem from Classical Gaelic (or Early Modern Irish), a conservative grapholect utilised from the end of the 12th century (Ó Cuív 1983: 3; McManus 1994) until the collapse of the Gaelic learned orders in the 17th (Ireland) and 18th (Scotland) centuries. Although Classical Gaelic was never institutionally imposed, its characteristics have informed literary standards into modern times.

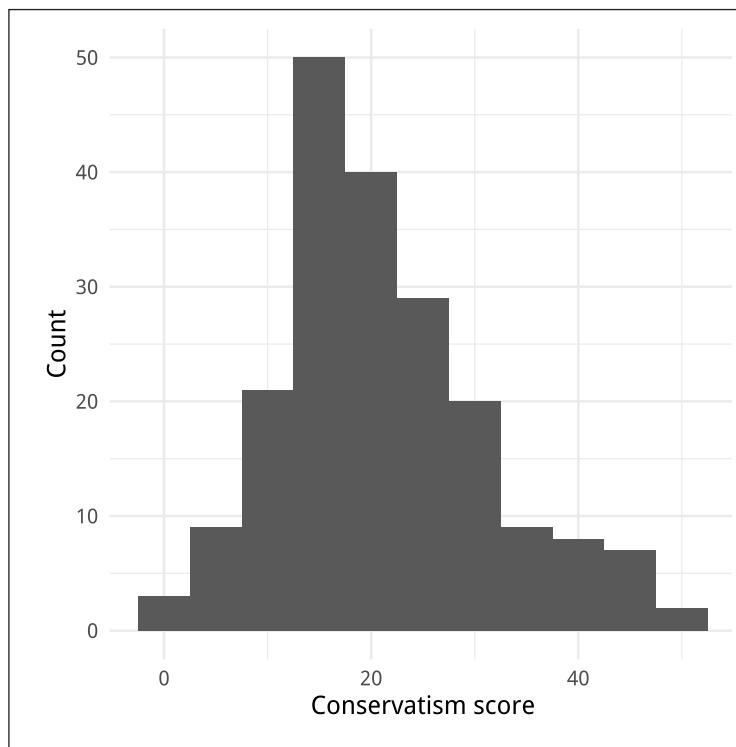
We derived an aggregate measure of how close a dialectal point is to a conservative maximum. Almost all of our features were coded so as to make 1 correspond to a return that agreed with the conservative norm and 0 to an innovative form. Therefore, an aggregate measure of conservatism can be derived by simply summing the responses for all features.⁸ Missing responses were treated as zero, as they also likely indicated a deviation from the conservative ideal.

As noted in Section 2.4, there were a total of 55 features; therefore, 55 is the maximum possible conservatism score. The descriptive statistics for the score are shown in Table 1. The maximum score found in the data is 50, at point 28 (Castlebay 1, Isle of Barra). Half of all observations lie between the values of 15 and 27. The distribution of the scores on a histogram with a bin width of 5 is shown in Figure 2.

⁸ One feature – an [z]ùil għlas had to be recoded, as voicing here is a innovation.

Table 1: Descriptive statistics for the conservatism score.

Measure	Value
Mean	21.17
SD	10.16
Minimum	1
Q1	15
Median	19
Q3	27
Maximum	50
(Theoretical max)	(55)

**Figure 2:** Distribution of the aggregate conservatism score.

It is notable that no point returned the maximum possible score of 55, corresponding to full agreement with the conservative ideal. The distribution of the scores in space is shown in Figure 3. The figure shows a pronounced east to west cline, from grammatically progressive in the east, to conservative in the west. The general pattern is interrupted occasionally by outliers, but we can explain at least some of these cases in terms of idiolectal variation. For instance, the contributor from Canna, one of the most conservative points, had a well-known *seanchaидh* (expert in oral tradition) for a father. Additionally, this individual's mother and husband were from Barra, the most morphologically conservative region in the LSS(G) materials. The informant from Dornie, in Kintail, was himself described as a *seanchaидh* with unusually archaic forms. Similarly, while the results for the Isle of Lewis diverge from the rest of the Outer Hebridean dialects and show similarities to the adjacent mainland in Wester Ross,⁹ point 3 (Bragar) stands out as particularly

⁹ This is not surprising, given the extensive historical ties between the regions; Lewis traditionally belonged to the county of Ross-shire, rather than to Inverness-shire as is the case with the rest of the Western Isles chain.

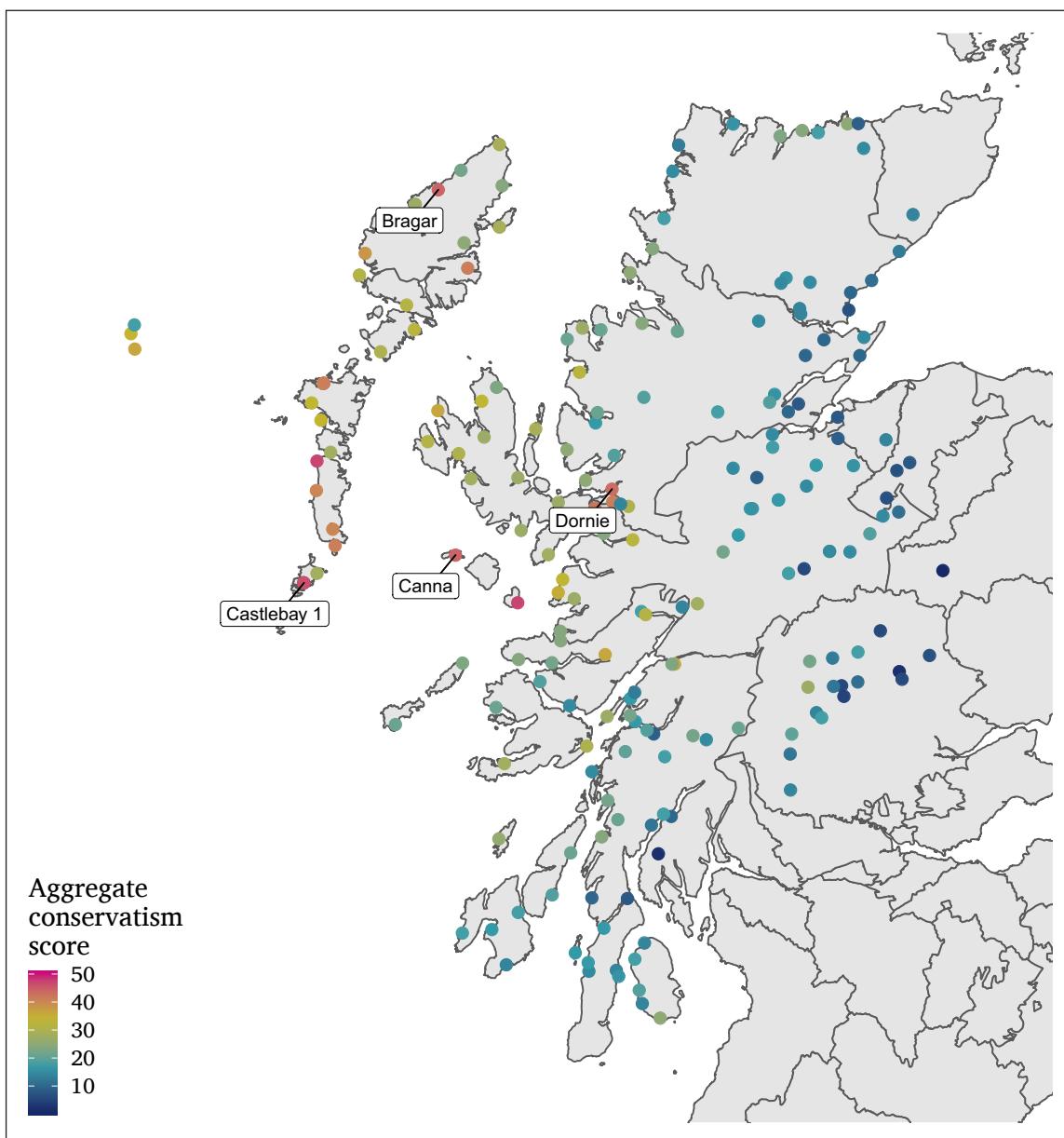


Figure 3: Distribution of the conservatism score.

conservative. Looking at Jackson's notes, we find that the contributor for Bragar was very literate and "perhaps too sophisticated to be an ideal informant".

3.2 Cluster analysis using edit distance

To identify whether the morphological variation observed in the data is congruent with the present state of knowledge on Gaelic dialect divisions, we subjected the data to a clustering analysis. In this section we report the results of this analysis when conducted on the raw questionnaire returns.

Since Gaelic orthography marks many of the relevant phonological and morphological distinctions, orthographic representations are suitable for dialectometrical methods using edit distance metrics such as Levenshtein distance (e.g. Nerbonne & Heeringa 2010). We converted the raw phonetic transcriptions into Gaelic spelling, normalising the otherwise often very narrow phonetic transcription of the questionnaire returns. We used a slightly modified version of Gaelic orthography, omitting some graphical devices that would artificially inflate the edit distances. In standard spelling, *t-sandhi* – the replacement of an

initial [s] by [t] in certain grammatical contexts – is expressed by <t-s>, as in *sùil* ‘eye’ [su:l] but (*leis an*) *t-sùil* ‘with the eye’ [tʰu:l]. The string distance between <s> and <t-s> is 2 (two insertions), although in reality we are dealing with a single edit (substitution of [t] for [s]). To avoid this, we retranscribed <t-s> as <t>.

As noted in Section 2.4, there were no returns for some of the prompts. Treating the missing data as empty strings produces artificially high distances: the edit distance between an empty string and a non-empty one equals the length of the non-empty string, producing on our data distances of 10 and more. To avoid these artefacts, we imputed the missing data by replacing blank returns with the mode (the most frequent response) for that prompt. The effect of this is to make dialects with missing points more similar to the “average” dialect, and reduce the likelihood of spurious outliers that might influence the clustering algorithm.¹⁰

We used the R package *stringdist* (van der Loo 2014) to calculate Levenshtein distances between the enquiry points for each of the prompts, and then summed the pairwise similarity scores for each prompt to arrive at a total similarity matrix. This was then subjected to a nested agglomerative clustering procedure using the Ward method, with the function *agnes* in the R package *cluster* (Maechler et al. 2015). We computed the *average silhouette width* for a number of partitions (between 2 and 10) to choose the best number of clusters. Silhouette width is a rough measure of how well each observation conforms with the other observations in the same cluster; in other words, partitions with a high average silhouette width mostly consist of internally coherent clusters, whilst low average silhouette widths signal that the partition does not do a particularly good job of capturing dissimilarities in the data. As a rule of thumb, average silhouette widths of <.2 can be interpreted as indicating lack of informativeness in a partition. On our data, the biggest number of clusters with an average silhouette width of >.2 was three. This clustering is shown in Figure 4.¹¹

The pattern that emerges from this exercise again shows a pronounced cline from the east to the west. The “eastern” cluster is concentrated on the mainland, covering basically all the points not on the western seaboard, in addition to the western and northern coasts of Sutherland, most of the western coast of Ross-shire and part of Lewis, as well as all of mainland Argyll south of Loch Linnhe and the islands of Arran and Islay. The “western” cluster includes most of the Inner Hebrides, Skye, and northern Argyll, and a few points in Lewis and Harris. Finally, points belonging to the “Hebridean” cluster are concentrated in the Outer Hebrides south of Harris.

At a very simplistic level, the existence of such a cline may indicate a relationship with the strength of the language in the respective communities. Indeed, the “Hebridean” cluster includes large parts of today’s “Gaelic heartland”, in areas such as Uist and Barra. On the other hand, points as far apart as Easter Ross, Central Perthshire, Kintyre, and Arran end up together in the same cluster, and they are all places where the regional variety of Gaelic is now extinct or moribund.

Nevertheless, this generalisation does not always hold up. It is true that the “eastern” cluster contains most of the points with the lowest percentages of Gaelic speakers. Yet, several points belonging to it did have high or very high percentages of Gaelic speakers in 1891, such as Lewis, Applecross and north-west Sutherland (the area known as

¹⁰ An alternative, suggested to us by a reviewer, would be to use the average distance for complete pairs only. However, in our data there are discrepancies in how much material is available for different points: missing data is not distributed uniformly. For this reason, the calculation of the averages would be based on very different sample sizes, making the comparisons less meaningful.

¹¹ See the supplementary material for the application of different methods to determine the best number of clusters. The results do not have a significant impact on our conclusions.

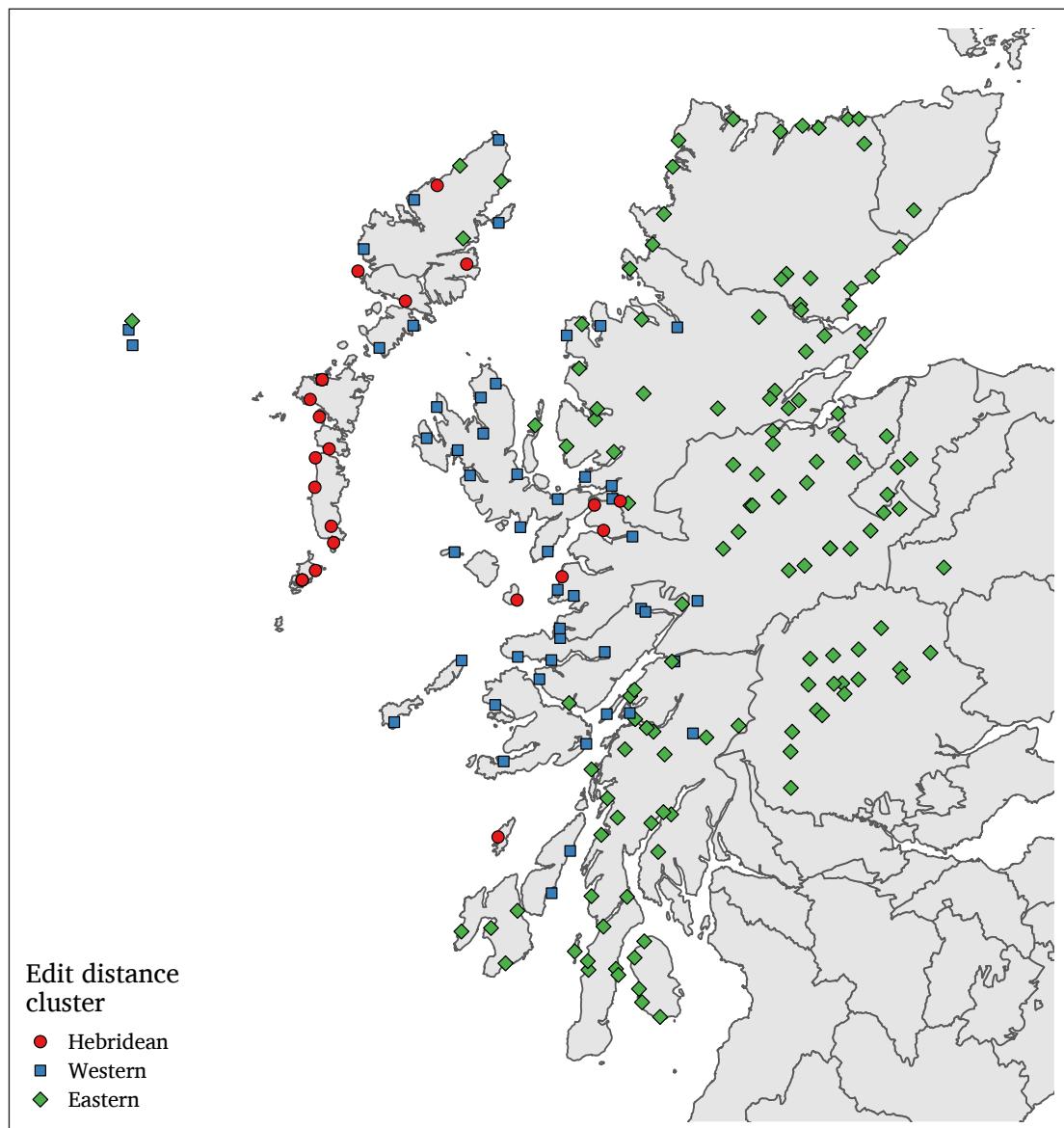


Figure 4: Clustering of dialects based on edit distance.

Strathnaver, or Mackay Country). Still, most of the points in the “Hebridean” cluster are near the top of the ranking of percentage of Gaelic speakers, whilst the “eastern” cluster consists mostly of points where about half the population was recorded as Gaelic speakers in 1891. Thus, we can say that the overall cline in this case is consistent with having some kind of relationship to language endangerment. Yet, in certain cases the correlation can break down, particularly in the north-west Gaelic-speaking area.

3.3 Cluster analysis of morphological features

Recall that the morphological data from the Survey materials were coded in a binary fashion, for the presence or absence of a particular feature. Since this coding treats the outcome variable as categorical rather than continuous, many clustering methods popular in the dialectometric literature (such as *k*-means clustering or agglomerative nesting) are not appropriate for our data. Instead, we conducted a *multiple correspondence analysis* (MCA). Multiple correspondence analysis (e. g. Husson, Lê & Pagès 2011) is a method for reducing the dimensionality of the data. It is similar in intent to methods such as principal components analysis or factor analysis, in that it produces a set of dimensions (smaller than the

original number of variables) and loadings for each data point in every dimension, so that the data points with a similar profile have similar loadings. Once the discrete data are represented as a set of continuous loading values, hierarchical clustering procedures can be applied; this method is known as *hierarchical clustering on principal components* (HCPC).

We conducted this analysis using the functions `mca` (for the multiple correspondence analysis) and `HCPC` (for the clustering) from the R package `FactoMineR` (Lê, Josse & Husson 2008). As with our conservatism score, we coded missing values (i.e. absence of a return) as negative returns, since either rejection of a conservative form or a missing return indicate that the expected conservative form is unlikely to be accepted by the informant.

The outputs of MCA are suitable for agglomerative clustering of the kind described in Section 3.2 above. In order to identify the best number of clusters in this case we used *inertia gain*, a measure widely accepted as appropriate for HCPC outcomes. Inertia is a measure of the heterogeneity of the data, both within and across the different clusters: if adding a partition leads to a significant drop in inertia, then adding such a partition is commonly considered meaningful, whereas partitions that only contribute modest drops in inertia are less likely to represent meaningful structure in the data. In our data, the best supported partition is into 4 clusters. Its spatial pattern is shown in Figure 5.

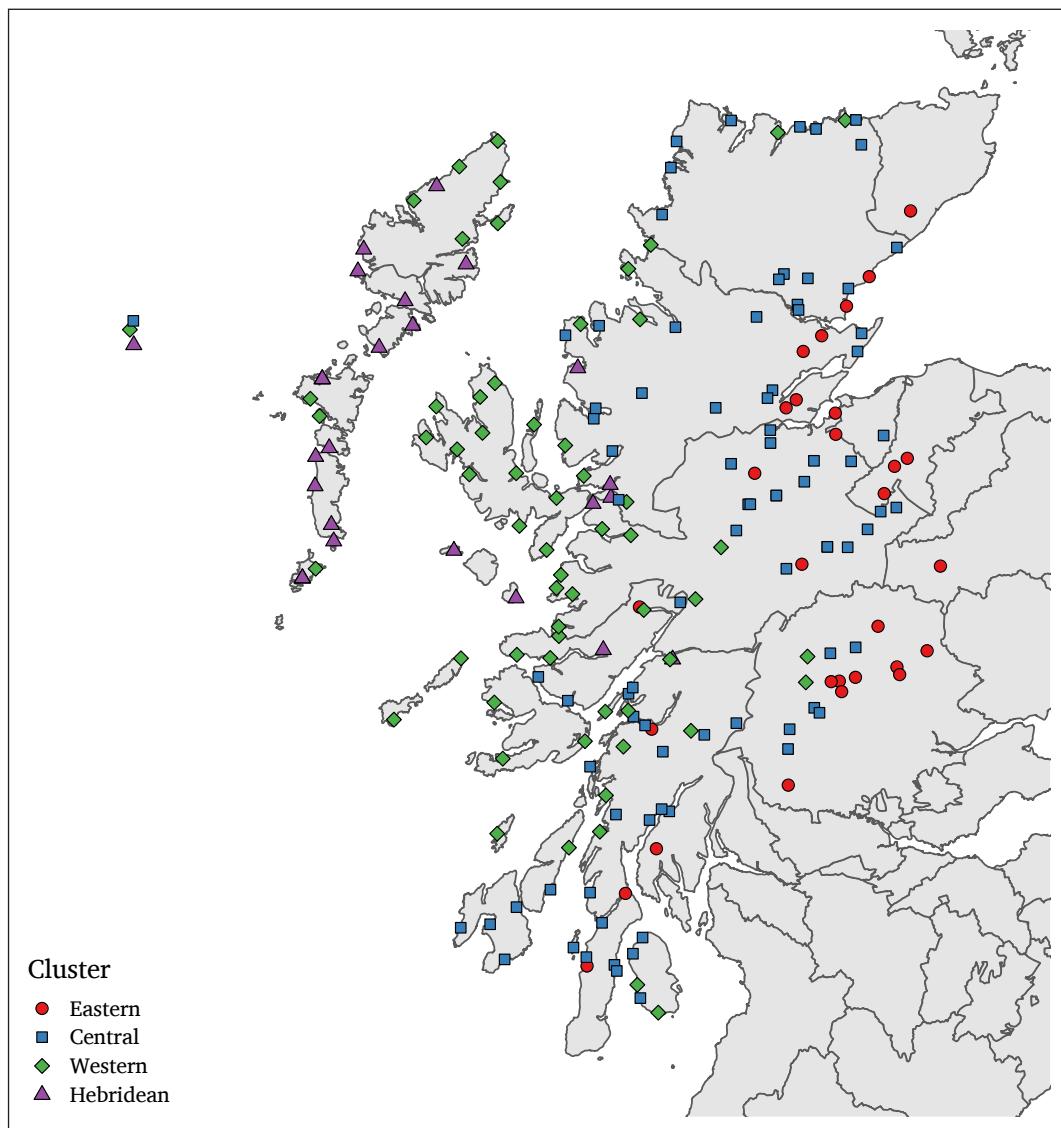


Figure 5: HCPC result with 4 clusters.

Once again, we observe an east-west cline. It is, if anything, even more pronounced than in the case of clustering based on edit distance. Thus, what we call the “eastern” cluster is concentrated along the eastern periphery in Perthshire, Nairnshire and the Ross-shire eastern seaboard. The next cluster, which we call “central”, is located further to the west on the mainland. It only occasionally touches the western seaboard, except in Argyll, but, remarkably, it includes both the southwestern-most points of the survey in Islay and far north-east of the Gaelic-speaking area in North Sutherland. A further one, the “western” cluster, covers most of the western seaboard, Mull, Skye, and (at least the northern part of) Lewis. Finally, the “Hebridean” cluster covers the area we identified in previous sections as relatively conservative, including most of the Hebridean chain from Harris southwards.

So far, the results of our study of morphology do not support the division of Gaelic dialects into a “centre”, covering the Hebrides and large parts of the western mainland, and a “periphery”, as offered by Jackson (1968), primarily on the basis of phonological criteria. What our results *are* reminiscent of is the dynamics of language endangerment, with a demographic “heartland” in areas such as the Outer Hebrides, Skye, and the western Highlands, and a much more patchy distribution of Gaelic speakers in the eastern and southern periphery, in places such as south Argyll, Perthshire, and the far north-east. The pronounced overall cline strongly suggests an explanation with roots in language attrition under conditions of progressive, subtractive bilingualism and the gradual weakening of Gaelic-speaking communities under the encroachment of English, with concomitant loss of morphological categories.

Cluster analysis does not really allow us to examine the potential role of demography and language endangerment in the development of morphosyntactic patterns. Pioneering work by scholars such as Dorian (1973; 1978a) has demonstrated the existence of this connection at a “local” level, i.e. within a single variety; however, the clustering analysis does not allow us to verify whether such a connection is also present across the board. To remedy this, we conducted a regression analysis to probe the connection.

3.4 Regression analysis: Conservatism and demography

The results of the cluster analysis described in the previous two sections allow us to discern an overall qualitative pattern of similarity between varieties. However, the cluster analysis is in many respects a crude tool; it does not facilitate identifying robust relationships between cluster membership and other potential properties of each data point.

To probe the relationship between our data and patterns of endangerment, we operationalised the state of the language in quantitative terms using census returns. We mapped the LSS enquiry points to census enquiry points and established the number and percentage of Gaelic speakers in each locality, using the census data collated by Duwe (2003). We utilised the 1891 census, as it was the closest to 1886, the average date of birth for LSS informants. In order to gauge the strength of Gaelic within the community, we used the percentage of Gaelic speakers at each point as our proxy for the state of the language at the time.

Apart from the demographic state of the language, a potential factor influencing the conservatism score is the sex of the informant. The Linguistic Survey of Scotland was conducted in a traditional dialectal framework targeting NORM informants, based on the assumption that males, at least in a traditional society, may be less open to linguistic changes (Chambers & Trudgill 1998: pp. 47, 61). Our results can be used to check that assumption.

Figure 6 shows the overall relationship between the share of Gaelic speakers in a locality and the aggregate conservatism score, by sex of the respondent, using a thin plate regression spline curve fit via the R package `mgcv` (Wood 2006). On average, as can be seen from the figure, this relationship is positive: a larger share of Gaelic speakers corresponds,

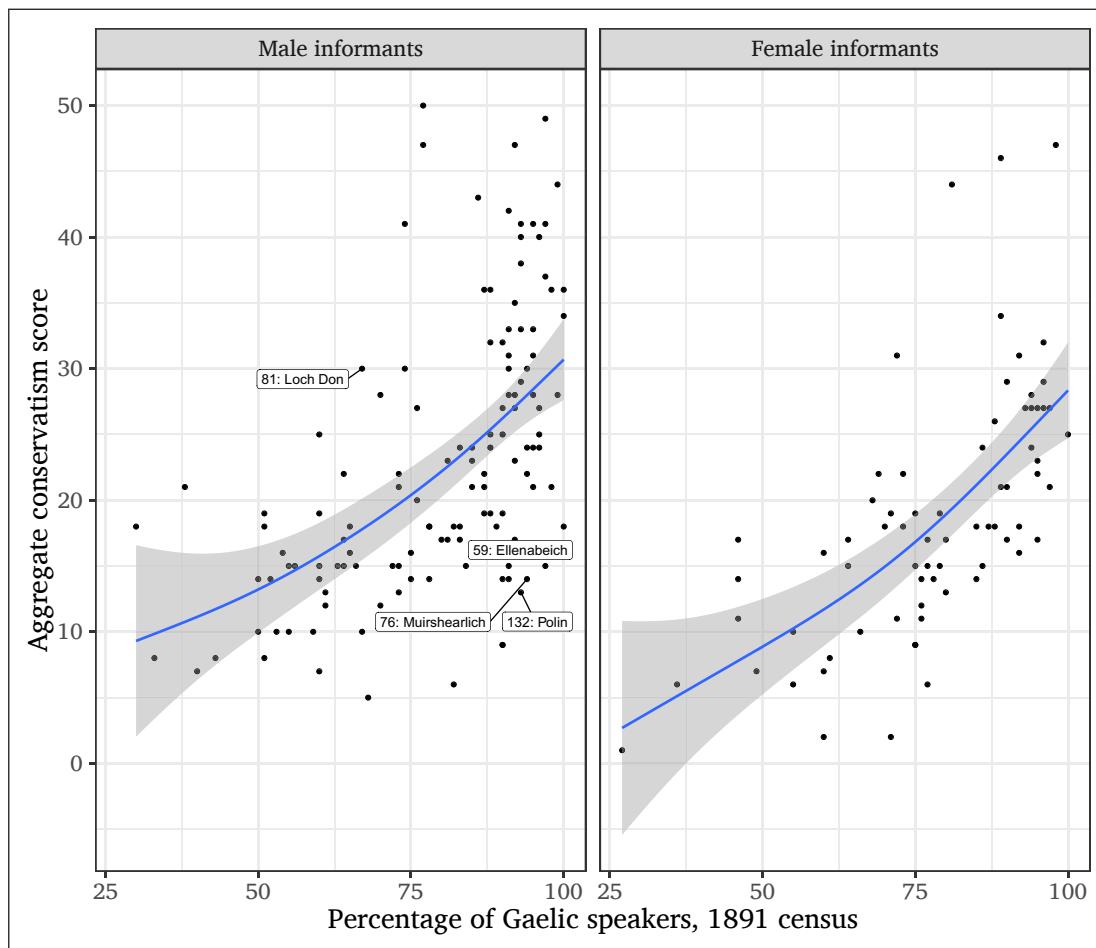


Figure 6: Overall relationship between demographics and the conservatism score.

overall, to a higher conservatism score. We can also see that the overall level of conservatism for male respondents is higher than for female respondents: this is confirmed by a two-sided t -test ($t(154) = 2.768, p = .006329$).

Nevertheless, we observe a significant number of outliers in both directions: we find points with a low percentage of Gaelic speakers and a high conservatism score, and vice versa. For instance, point 81 Loch Don (Isle of Mull) is in the lowest quartile for percentage of Gaelic speakers, but in the highest quartile for conservatism; conversely, points 59 Ellenabeich (Isle of Seil), 76 Muirshearlich (in Lochaber, near Fort William), and 132 Polin (Assynt) are in the highest quartile for speaker percentage but in the lowest quartile for conservatism.

In order to further probe the relationship between informant sex, date of birth, demography, and space, we fit a range of non-linear regression models with conservatism as the dependent variable. We used generalised additive models implemented in the R package `mgcv` (Wood 2006), treating the share of Gaelic speakers and year of birth as smooths to account for any possible non-linearity in the effect. We also introduced a group-level coefficient (“random effect”, in this case an intercept) of registration district. The reasoning behind the introduction of district was that it could be used as a proxy for geographical proximity, allowing us to check for any spatial non-heterogeneity.

The best model is summarised in Table 2. Model comparison was conducted using likelihood ratio tests and the Akaike Information Criterion (more specifically second-order AIC, or AIC_c , as recommended by Burnham & Anderson 2004 for sample sizes as small as ours. For details on the model selection procedure and test results, see the R notebook in

Table 2: Summary of the regression model.

Parametric coefficients	Estimate	Standard error	t-value	p-value
Intercept	23.3060	1.1999	19.4235	<0.0001
Sex: female	-2.0465	0.9508	-2.1523	0.0329
Smooth terms	Effective df	Reference df	F-value	p-value
Percentage of Gaelic speakers	0.9129	4.0000	54.9925	0.0009
Random intercept: district	45.5267	57.0000	4.1020	<0.0001

the Supplementary Material). In this model, the overall effect of demography is positive, in that larger percentages of Gaelic speakers result in a higher conservatism score. The district random effect is also included in the model, and it accounts for spatial heterogeneity. Specifically, models without this effect showed significant spatial autocorrelation in the distribution of residuals; that is to say, in models without the random effect of district there was spatial variation in how well the main effects were able to predict the conservatism at each point. In the best model, which does include the random effect of district, the distribution of residuals is much closer to the normal, and the residuals also showed a random spatial distribution.¹²

Figure 7 visualizes this spatial variation by district. It shows the LSS(G) points with the estimated random effect for the district to which the point belongs (see the R notebook for estimates of effects and confidence intervals). We can still observe an east-west cline, but the picture is less noisy than one based on raw data. This is partly because we have been able to control for some confounding variables, and partly because the grouping of points into districts is, by necessity, a pooling exercise. The clearer picture emerging in Figure 7 appears to indicate real diatopic variation, by exposing divisions that were less clear in the raw data: see in particular the visible differences between the northern and the southern parts of the Outer Hebridean chain, and the north-west mainland.

To summarise our spatial analysis of the aggregate conservatism score, we observe a pronounced east-west cline in our data, whereby the western areas preserve the “Biblical Gaelic” morphological “ideal” to a greater extent than eastern ones. This distribution, perhaps unsurprisingly, shows a non-trivial relationship with the dynamics of language endangerment and language shift. Nevertheless, the relationship between the demographic context and morphological conservatism is not straightforward: while some areas represent the overall relationship between these two factors, other zones show greater or lesser conservatism than expected on the basis of the full data. Thus, there appears to be an interaction between geographic space and the speed with which language endangerment influences morphological developments, with different varieties implementing this relationship differently. In other words, language shift alone appears to be insufficient to explain the morphological effects visible in the overall east-west conservatism cline, and the data reveals a genuine diatopic signal.

3.5 Correlation analysis: Features

In this section, we show how quantitative analysis of survey data may allow us not only to answer specific questions related to the properties of individual Gaelic varieties and their place within the wider dialectological picture, but also to give insights into the underlying grammatical structure of the language. The morphological survey data is quite “shallow”,

¹² A Geary’s test calculated using the R package `spdep` (Bivand & Piras 2015) shows a C statistic of 0.9999, with a standard deviate of -0.0012 and a p value of 0.4995, indicating lack of spatial autocorrelation (see the R notebook for details).

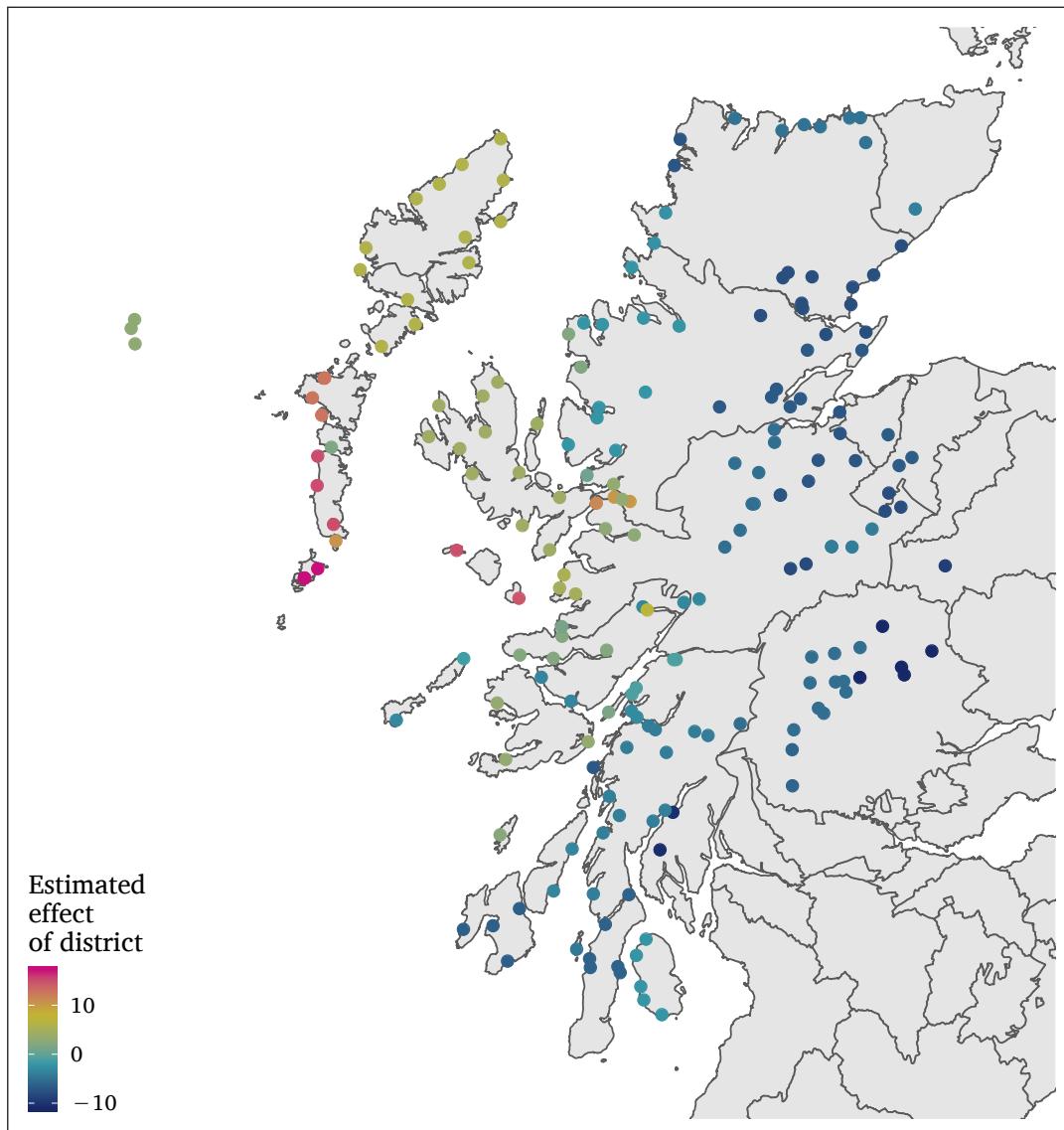


Figure 7: Estimated district random effect at each point.

focusing on lexical allomorphy and surface morphophonology such as consonant mutation or vowel changes. It cannot be taken for granted that similar surface phenomena necessarily reflect identical structural processes (e.g. lenition across different grammatical contexts). Conversely, it is possible that the same structural process could be responsible for more than one surface phenomenon (e.g. gender agreement influencing a variety of surface phenomena). The latter point is strongly emphasised in Adger's (2017) study of the syntactic mechanisms active in Gaelic language obsolescence. Adger argues that a single syntactic change – the loss of agreement features – produces visible effects in an entire set of apparently unrelated surface constructions, from the expression of possession, to the syntax of pronominal objects of non-finite verbs, to the structure of passive constructions. His argument focuses on the changes apparent under conditions of obsolescence in the East Sutherland variety described by Dorian (1978a): he suggests that the co-occurrence of all these “surface” changes in that variety reflects a single modification of the underlying grammar.

Here, we extend this thinking to a broader range of observed features. Following Adger (2017), we suggest that co-variation of two or more surface features deserves to be taken seriously as a potential indicator of the commonality of the grammatical process

underlying them. The logic is simple: if grammatical change X leads to surface changes Y and Z, then varieties undergoing X should show both Y and Z, too.

To evaluate this proposition quantitatively, we conducted a correlation analysis of the features in the dataset. Once again, we treated missing returns as negative responses, to reflect the fact that both a lack of return and a negative response indicate non-adherence to the conservative ideal. Following this transformation, we calculated the correlation between every pair of features and tested the significance of each correlation coefficient. The correlation was calculated using Spearman's ρ criterion, which is appropriate for our categorical data, since it is computed using rank order rather than treating 1 and 0 as values of a continuous variable.

The results of the calculations are presented in Figure 8. The figure presents the correlation matrix in graphical form. Only those cells where the correlation coefficient was significant at $p < .01$ are shown in the plot. The size of the circle indicates the absolute value of the correlation coefficient, as does the depth of the colour. Red hues indicate positive correlations, and blue hues indicate negative correlations. The features are ordered in the matrix based on the outcomes of a Ward agglomerative clustering procedure and the HCPC (hierarchical clustering on principal components) methodology. That clustering is also reflected in the rectangles, which delimit 20 clusters. This is done simply to better visualise which of the adjacent features do cluster together; no significance should be attached to the number of clusters.

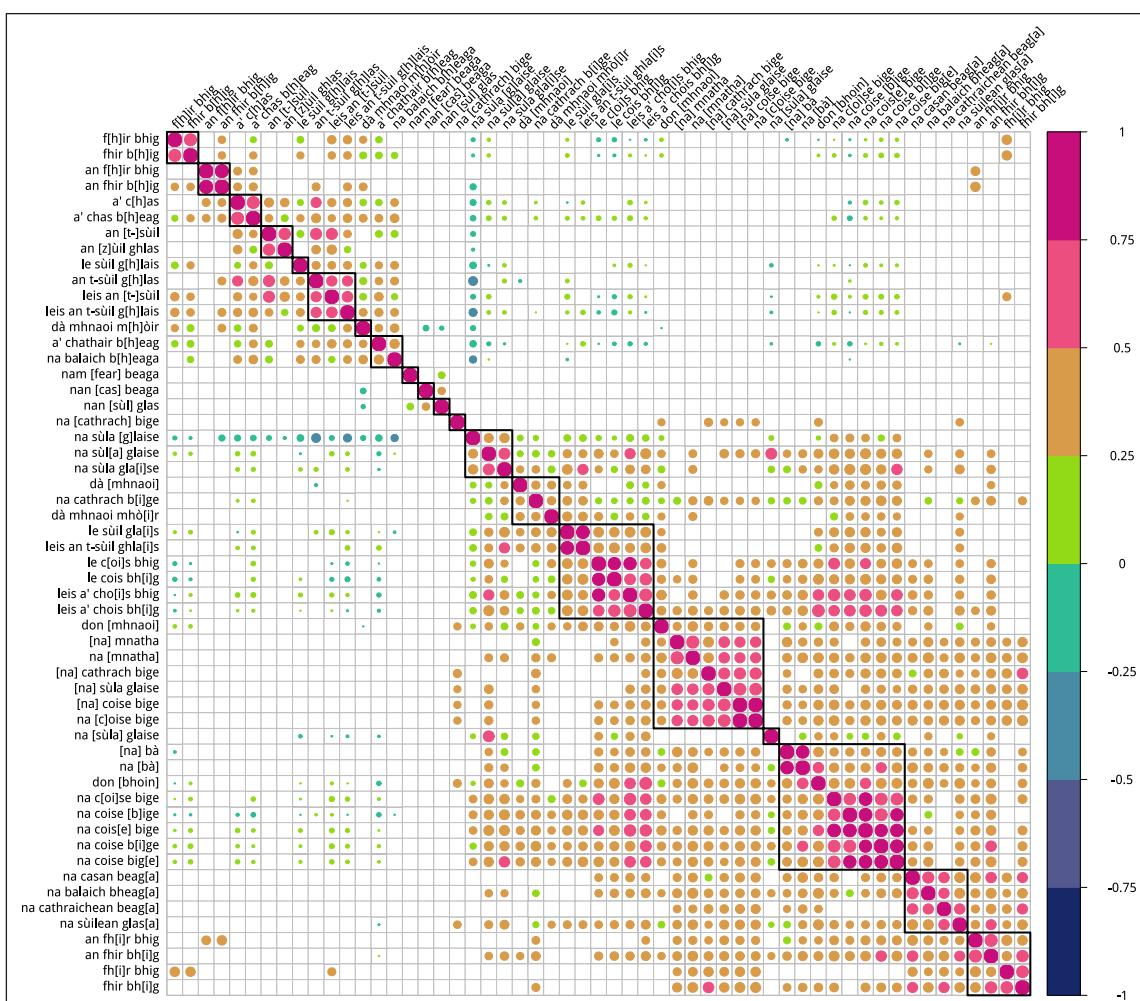


Figure 8: Correlation plot of features.

This analysis succeeds in bringing out the internal structure of the data. Many of the features in the questionnaire correspond to a single grammatical phenomenon elicited in multiple contexts. For instance, we can observe that there is a strong correlation between the returns regarding the presence of a plural suffix in monosyllabic adjectives: the features *na casan beag* [a] (after a feminine 2nd declension noun), *na balaich bheag* [a] (after a masculine 1st declension noun), *na cathraichean beag* [a] (after a feminine 5th declension noun), and *na sùilean glas* [a] (after a feminine 3rd declension noun) are well correlated. This is expected, since the presence of this suffix should depend only on the number of syllables in the adjective itself, rather than the gender or declension class of the preceding noun.

That the algorithm is able to recover such hidden structure can serve as a sanity check to give us some confidence that the correlations it uncovers correspond to grammatical or lexical categories. Having established this, we can now proceed to examine several feature bundles more closely.

3.5.1 Dative slenderisation: Erosion of morphophonology

In this section, we focus on several clustered features that show the presence of slenderisation (Section 2.2) in dative contexts, affecting 2nd declension nouns and monosyllabic adjectives. The relevant features are *le sùil ghla[i]s*, *leis an t-sùil ghla[i]s* (showing adjectives after both indefinite and definite 3rd declension nouns), *le c[oi]s bhig*, *leis a' ch[oi]s bhig* (indefinite and definite 2nd declension nouns), and *le cois bh[i]g* and *leis a' chois bh[i]g* (adjectives after indefinite and definite 2nd declension nouns). As we can see, slenderisation is a pattern present across declension classes and definite and indefinite contexts, just as was the case with plural adjective suffixes in the previous section.

In progressive grammar, slenderisation is typically not found in the dative. The pattern cannot be interpreted, however, as reflecting across-the-board loss of slenderisation, since the featural cluster is not particularly well correlated with other features that also demonstrate slenderisation: for instance, there are not close correlations with features showing slenderisation in the genitive singular of 1st declension masculine nouns and adjectives (*an fh[i]r bh[i]g*), or in the vocative (*fh[i]r bh[i]g*). Similarly, dative slenderisation does not show a close relationship with other features associated with the dative case in our dataset, such as *t*-sandhi.

This result suggests that slenderisation in the dative is a single phenomenon (since it behaves consistently across parts of speech and declension classes), but its loss does not reflect either the loss of slenderisation from all contexts or the loss of the dative case from the grammar. Consider again the lexical item for the dative singular given in (2) and repeated here for convenience:

$$(3) \quad [\text{DAT SG}]: \left\{ \begin{array}{l} [\text{FEM}] \Leftrightarrow ^j \\ \Leftrightarrow \emptyset \end{array} \right\}$$

If slenderisation were being lost across the board, it could be formalised as the loss of the floating palatalising feature $[^j]$ from lexical items where it is present. This cannot be the case, however, as changes in slenderisation affect different contexts differently. If it were the dative case itself that had become obsolete, the process could be formalised as the loss of all lexical items expressing the features [DAT SG]. Yet, again, this cannot be the case, since such items are necessary to account for the feature of *t*-sandhi, which behaves differently from the dative slenderisation features. We conclude, then, that the change in question affects the subcategorisation conditions in the lexical entry in (3): specifically,

the [FEM] clause is lost, leaving only the “elsewhere” zero exponent. This gives the correct result, namely loss of slenderisation in both nouns and adjectives, but only in the dative singular. Note that this pattern cannot be due to a failure of agreement, i.e. the lack of a transfer of the [FEM] feature from the noun to the adjective: if the lexical item were kept as in (3), this would have led to slenderisation persisting in nouns.

The analysis in this section suggests that it is fruitful to consider the underlying grammatical categories and their morphological and morphophonological exponence separately if we are to ascertain what processes are involved in the changes documented in our data set. In the following section we use this result to consider the patterning of grammatical categories associated with *multiple* exponents in Scottish Gaelic.

3.5.2 Lenition and slenderisation

Lenition in the vocative singular of 1st declension masculine nouns, like *fear* ‘man’ (feature $f[h]ir\ bhig$), is strongly correlated with lenition in monosyllabic masculine singular adjectives, like *beag* ‘small’ (feature $fhir\ b[h]ig$). This suggests that lenition in both nouns and adjectives in the vocative is driven by a single grammatical mechanism. We argue, however, that changes in the pattern of lenition in these cases do not necessarily indicate a change in the existence of the vocative as a grammatical category.

This is so because the vocative in these morphological classes exemplifies multiple exponence: it is expressed both by lenition and by slenderisation: the standard vocative form of *fear* [fer] is *fhir* [ir^j]. The features for slenderisation in these contexts ($fh[i]r\ bhig$ for the noun and $fhir\ bh[i]g$ for the adjective) do not pattern particularly strongly with the lenition features. (They do, however, cluster with each other.) We conclude, therefore, that masculine 1st declension nouns like *fear*, and masculine adjectives like *beag* use the same morphophonological mechanisms to expon the grammatical category VOCATIVE SINGULAR: specifically, in both cases the exponents contain the floating palatalization feature /j/ at the right edge and whatever manner feature triggers lenition at the left edge. Adjectives and nouns then pattern together because the [VOC SG] feature(s) are involved in morphosyntactic agreement mechanisms, ensuring that these exponents are present in adjectives and nouns simultaneously. However, language change can target these two floating features individually: a change in exponence removing one of these phonological items does not necessarily result in the erosion of the grammatical category itself. Much the same conclusion applies to the use of lenition and slenderisation in the same two morphological classes in the genitive singular.

Changes in the behaviour of lenition and slenderisation in these particular categories, however, do not necessarily correlate very strongly with the behaviour of these morphophonological processes in other grammatical contexts: as we saw, slenderisation is also used in the dative case, but the attrition of that mechanism does not seem to be closely associated with its attrition in the genitive and vocative. We conclude, therefore, that the behaviour of multiple exponents in the genitive and vocative case is consistent with our approach to feminine dative slenderisation: the change observed in our data targets individual aspects (lenition or slenderisation) of lexical entries separately, but change in exponence does not necessarily imply change in the underlying grammatical specification.

In the last two sections, we have considered cases where the same means of exponence is observed in multiple grammatical contexts, and shown that language change can affect an exponent in one context without necessarily leading to a loss of the underlying grammatical mechanisms. In the following sections we demonstrate how our data set can be used to uncover changes that are different on the surface but do reflect a single underlying change in the grammar.

3.5.3 Gender and mutation

The expression of morphological gender in Scottish Gaelic, as in the Gaelic languages more generally (see Frenda 2011 on the closely related Irish), is multifaceted, but primarily concerns initial mutations and agreement with the forms of the article. Specifically, most consonant-initial feminine singular nouns undergo lenition (or *t-sandhi* in certain cases) if they are preceded by the definite article, and trigger lenition themselves on a following adjective. The definite article also exhibits agreement, for example, in the genitive singular, where the form is *a' (an)*, with lenition, before masculine nouns, and *na*, without lenition, before feminine nouns. Our dataset allows us to probe both of these means of exponence.¹³

We can see that certain exponents of gender cluster together very strongly. This is particularly the case for lenition in the nominative singular, which applies to feminine nouns after the article (*a' c [h]as*) and to adjectives after the nouns (*a' chas b[h]eag*). Similarly, the features describing the shape of the article for the genitive singular before feminine nouns of different declension classes ([*na*] *mnàtha*, [*na*] *cathrach bige*, [*na*] *sùla glaise*, [*na*] *coise bige*) also cluster with each other.¹⁴ Crucially, however, the patterning of gender-driven mutation and the patterning of agreement with the definite article are quite different.

Specifically, the loss of gender agreement between the noun and the article in innovative genitive forms such as *a' chas bheag* (for conservative *na coise bige*) does not imply the loss of gender distinctions more generally. This is because gender distinctions can be preserved in other contexts: for instance, feminine nominative forms like *a' chas bheag* contrast in terms of lenition behaviour with masculine nominative forms like *am fear mòr* ‘the big man’. In principle, the existence of innovative genitive forms like *a' chas bheag* that are identical to the nominative could be ascribed to a loss of the grammatical contrast between nominative and genitive more generally. However, our data does not show a strong correlation between the “feminine article” features and the features related to segmental exponence of the genitive in masculine nouns (such as *an fh[i]r bhig*): even if feminine nouns end up not contrasting genitive and nominative forms, masculine nouns can still preserve that distinction.

Our conclusion, therefore, is that the attrition of agreement between the noun and the article, i.e. the use of *a(n)* as the genitive singular article with (historically) feminine nouns instead of *na*, does not imply the loss of gender specification of the noun itself. The cross-dialectal data indicates that attrition can target the noun-article agreement alone, rather than the gender system in its entirety. Notably, the weakening of agreement feature specifications, which would be necessary to implement the noun-article agreement, is exactly the syntactic change observed by Adger (2017) on the basis of different data. Our suggestion here is that what is lost is specifically agreement between the noun and the determiner for the gender feature.¹⁵ Differences in changes affecting different exponents of gender are also observed by Frenda (2011) in Irish, where he finds a difference in the patterning of article-related agreement and agreement with pronominal referents.

However, the exponence of gender also provides examples of changes in morphophonological patterns being driven by changes in the underlying grammar. This comes across most clearly in the clustering of the features *a' c [h]as bheag* and *a' chas b[h]eag*,

¹³ Gender is certainly relevant for pronominal reference and various processes in syntax, as Adger (2017) discusses in detail, but this data is not available to us here.

¹⁴ The exception here is [*na*] *bà*; however, this may have more to do with the reassignment of the noun *bò* into the default class of masculine 5th declension nouns, or its substitution by a different lexical item, such as the (masculine) *màrt*.

¹⁵ A reviewer suggests an alternative analysis where the [FEM] feature is deleted from the noun in some contexts (such as in the presence of definiteness). We leave the theoretical exploration of these results for further research.

indicating the presence of lenition in nominative feminine singular nouns after the definite article, and in adjectives following singular feminine nouns, respectively. As we saw in the previous sections, it should not at all be impossible for attrition to target the exponent (here: lenition) only in certain morphological contexts, for instance for lenition to be lost only on nouns (giving a return like *an cas bheag*) or only on adjectives (giving *a' chas beag*). However, this is not what we observe: instead, nouns and adjectives appear to pattern together. They also pattern closely together with one other feature involved in the morphology of gender, but in a different lexical item, namely *t-sandhi* after the definite article (feature *an [t-]sùil*). That feature, in turn, also correlates with its adjectival counterpart *an t-sùil g[h]las* (although the clustering algorithm does not put them quite as close to each other as in the case of *cas*).

We suggest that these patterns indicate that lenition in the feminine singular is driven by the same underlying grammatical mechanism in nouns and in adjectives (just as, say, lenition in the 1st declension vocative singular discussed in the previous section), and there is some evidence that attrition targets these features across lexical items, since all of the relevant features pattern together quite closely.¹⁶ Thus, the pattern in these cases indicates that language change targets not specific mechanisms of exponence, as exemplified in Sections 3.5.1 and 3.5.2, but the underlying grammatical phenomenon, namely feminine gender: if gender features are absent, then lenition is lost simultaneously both in nouns after the article and in adjectives after nouns. In this respect, our results reproduce the findings of Adger (2017) within a dialectometric context: we see how the same grammatical phenomenon – loss of gender features – has recurring effects on disparate surface phenomena not only within a single variety (as shown by Adger 2017) but also across varieties.

3.6 Discussion

As reported above, this work has two complementary strands: an aggregate analysis of diatopic morphological variation in Scottish Gaelic, which we have conducted partly as a methodological sanity check, and an attempt to infer deeper grammatical patterns through examining the behaviour of morphological features across varieties.

To gauge the patterns of diatopic variation, we used well-established dialectometrical methods such as edit distance, agglomerative clustering and multiple factor analysis. We also introduced a measure of morphological conservatism to examine the degree of individual survey points' adherence to the established conservative norm. On the basis of this data, the Outer Hebrides – especially the southern isles such as the Uists and Barra – emerged as the most conservative region. Furthermore, a relatively strong east-to-west cline obtained, with innovative forms being increasingly more common towards the eastern periphery.

Although these findings confirm some modern lay expectations, in many ways they disagree with Jackson's (1968) postulation of a more conservative “centre” and a more progressive and fragmented “periphery”. Rather, what we see here is relatively clear differentiation between the mainland – apart from isolated areas on the west coast and Lewis – and the more conservative Hebridean region. We also examined the hypothesis that the

¹⁶ The exception here is the feature *a' chathair b[h]eag*, which shows lenition of the adjective after a 5th declension feminine noun. However, this is an artefact of the data: unlike most other declensions, the prompts for the 5th declension ended up using a variety of head nouns, depending on what items were known to the informant: in the raw returns, we find *cathair* ‘chair’, *nathair* ‘snake’, *iuchair* ‘key’, *uair* ‘hour’, and *acair* ‘anchor’ (and also *séithear*, a borrowing from English *chair*). Sometimes the particular noun elicited may have been reassigned to the masculine gender, in which case the lack of lenition is in line with the conservative system and does not necessarily indicate a full breakdown of the gender system.

patterns of variation observed in the data are due to language attrition and obsolescence in the context of language minoritisation. Although it appears that language decline does play a role in the maintenance of conservative norms, a clear diatopic signal is also present in the data, undermining the opinion, previously expressed by some authorities, that Gaelic morphosyntax does not exhibit significant variation in space.

The results discussed in Section 3.5 extend the findings of Dorian (1973; 1977) and Adger (2017) beyond a single variety. Dialectometrical methods allow us to quantify whether the effects observed “up close” in the examination of a single dialect are found across a larger area, and hence whether the results can be said to hold true for the language at large.

A particular advantage of a quantitative approach is that it can leverage hidden structure within the data. Clustering algorithms consider not just the pairwise relationship between the features of interest, but also whether they are similar in how they relate to other features within the dataset. If several surface features reflect the same grammatical phenomenon, then it is straightforwardly predicted that these features should co-occur whenever the underlying grammatical phenomenon is present. A second, more subtle prediction, is that these co-occurring features should have a similar distribution vis-à-vis *other* features, again because their distribution should reflect the distribution of the underlying phenomenon. This is something that is difficult to verify within a single variety, but dialectometry provides a way of achieving this result.

In our dataset, these methods have allowed us to establish quite firmly the difference in behaviour between morphophonological exponents (e.g. consonant mutation and slenderisation), their behaviour in the context of a particular morphological category, and the morphological categories themselves. As our results show, we can in principle distinguish between three types of change:

- Wholesale loss of an exponent. We do not observe this with morphophonological phenomena such as mutation, but the patterning of the adjective plural suffix *-a* discussed in Section 3.5 can be interpreted in this way;
- Loss of the use of a particular exponent for a morphological category, without the loss of the category itself. We saw this with the loss of dative slenderisation without the loss of dative (Section 3.5.1), and the separate behaviour of lenition and slenderisation in the vocative (Section 3.5.2);
- Loss of the morphological category itself, which leads to the loss of all surface features associated with it. This, as we argued, is the case for grammatical gender: its loss leads to the loss of mutation in nouns after the article and adjectives after nouns.

The distinction between exponents and categories that we have pinpointed corresponds to that made by Dorian (1973; 1977) in her study of initial mutations in East Sutherland Gaelic. She distinguishes between having access to “a repertory of initial consonant choices” (including lenition and nasalisation), and the use of those consonant choices in different grammatical contexts.¹⁷

The findings underline the importance of considering change in the exponence of grammatical categories separately from change in the behaviour of the grammatical category itself. This is particularly important given the socioeconomic context of Gaelic as a minoritised, endangered language. As we have repeatedly noted, many of the changes observed in the language have been ascribed to language obsolescence (notably by Dorian 1973; 1978b, but contrast Hamp 1989). Non-compliance with the prescribed conservative ideal

¹⁷ In Dorian (1973), no speaker has entirely lost access to the mutation system: in some contexts, all or almost all speakers (albeit not necessarily semi-speakers) continue to apply mutation rules.

is, of course, also commonly interpreted in this context in lay discourses (Bell et al. 2014). However, as we have seen, the loss of a particular exponent such as consonant mutation or slenderisation does not necessarily imply the breakdown or loss of the underlying element of the grammatical system.

To summarise, then, our study has demonstrated how dialectometry can be useful not only for the study of diatopic, social, or diachronic variation within a set of varieties, but also in order to discover the underlying grammatical structure, in ways that are less accessible to traditional methods.

4 Conclusion

Our aims in this paper have been threefold. First, we have used dialectometric methods to illuminate various aspects of diatopic variation in Scottish Gaelic. We have demonstrated the existence of significant dialectal variation in the language in an area – morphosyntax – where it has not been previously studied. We also have established the patterns of this variation and explored their relationship with other factors, notably language endangerment. Apart from the intrinsic value our results have for understanding spatial variation within the Gaelic-speaking area, they also have important implications for language planning. As we have seen, no informants in the LSS(G) materials exhibited a system retaining all of the grammatical features of the received conservative standard, even in the relatively formal context of a dialectological interview. In fact, given the median conservatism score of 19 (out of 55), it can be argued that when the LSS(G) materials were gathered in the mid-20th century, a significant proportion of Gaelic speakers used a system that deviated from the accepted conservative standard. This has clear implications for corpus planning in the present context, given the symbolic value ascribed by the Gaelic-speaking community to the norms prevailing at that time (Bell et al. 2014).

Second, in addition to its contribution to Scottish Gaelic studies, the paper has also aimed to make a methodological contribution. It has further demonstrated the viability of dialectometry, extensively used for the study of phonological variation, to morphosyntax, following recent work by authors such as Spruit (2008); Heeringa et al. (2009); Szmrecsanyi & Kortmann (2009); Szmrecsanyi (2013); Heeringa & Hinskens (2014); Scherrer & Stoeckle (2016); Aurrekoetxea (2016), and it is also the first dialectometric study of the morphosyntax of a Celtic language.

Third, and finally, we have shown that dialectometrical findings may be of interest also to theoretical linguists. Quantitative studies such as ours make it possible to generalise about the behaviour of individual features beyond a single variety of the language. More importantly, these methods have allowed us to uncover commonalities in the behaviour of morphosyntactic features (Adger 2017) that are not always apparent from the examination of individual varieties, and make inferences about the underlying grammatical structure. Thus, dialectological data – whether from existing traditional dialect surveys or newly gathered – can also provide interesting material for theoretical enquiry.

This study, of course, has a number of limitations. The underlying data is, in many respects, noisy, due to the nature of traditional dialectological work. There are also methodological imperfections in the original collection of data, biases in informant selection, and so on. In any case, we have only considered a selection of the morphosyntactic material available, and in the future it will be important to consider the full available data. An important future prospect for the study of Scottish Gaelic is combining the morphosyntactic results with the wealth of phonetic data, for instance as presented in the published *Survey of the Gaelic Dialects of Scotland* (Ó Dochartaigh 1994–1997). Further refinements to the methods are also possible, such as more detailed coding of the returns, improved methods of calculating string distances (for instance, using pointwise mutual information Levenshtein

distance; Wieling, Prokić & Nerbonne 2009) and of aggregate conservatism measures, or different clustering algorithms. Nevertheless, we suggest that this paper provides a successful proof of concept for theoretically informed dialectometrical research into morphosyntactic variation.

Abbreviations

DAT = dative, FEM = feminine, GEN = genitive, HCPC = hierarchical clustering on principal components, LSS(G) = Linguistic Survey of Scotland (Gaelic), M = masculine, MCA = multiple correspondence analysis, NOM = nominative, NORM = non-mobile older rural males, PL = plural, SD = standard deviation, SG = singular, SGDS = Survey of the Gaelic Dialects of Scotland, VOC = vocative

Supplementary Files

The Appendices referred to in the text are contained within the Supplementary Material, which is available at Open Science Framework at <https://osf.io/gbewz/>. The raw returns (coded as described in Section 2.4), coded data and an R notebook reproducing the analysis in the paper (including additional discussion as indicated in the paper) are available at Open Science Framework as Iosad & Lamb (2020).

Acknowledgements

The authors would like to thank the staff of the School of Scottish Studies Archives at the University of Edinburgh, particularly Cathlin Macaulay and Caroline Milligan, for help with access to the LSS(G) archival materials, and for permission to release the raw data under a CC-BY license. For useful feedback and suggestions, thanks are due to the audience at the Rannsachadh na Gàidhlig 2016 conference (Sabhal Mòr Ostaig UHI), where an early version of this paper was presented, and to Warren Maguire and Wilson McLeod; at *Glossa*, we received very valuable feedback from three anonymous reviewers and the editor, Johan Rooryck.

The maps were created using a shapefile with Scottish county boundaries provided by A Vision of Britain. This work is based on data provided through <https://www.VisionofBritain.org.uk> and uses historical material which is copyright of the Great Britain Historical GIS Project and the University of Portsmouth. This work is made available under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported Licence.

Competing Interests

The authors have no competing interests to declare.

References

- Adger, David. 2017. Structure, use, and syntactic ecology in language obsolescence. *Canadian Journal of Linguistics* 62(4). 614–638. DOI: <https://doi.org/10.1017/cnj.2017.32>
- Adger, David & Gillian Ramchand. 2006. Dialect variation in Gaelic relative clauses. In James E. Fraser, Anja Gunderloch & Wilson McLeod (eds.), *Rannsachadh na Gàidhlig* 3, 179–192. Edinburgh: Dunedin Academic Press.
- Anderson, Cormac. 2016. *Consonant colour and vocalism in the history of Irish*. Poznań: Adam Mickiewicz University dissertation.
- Aurrekoetxea, Gotzon. 2016. Analysis of the morphological variation of Basque. *Dialectologia et Geolinguistica* 24. 21–41. DOI: <https://doi.org/10.1515/dialect-2016-0002>
- Bell, Susan, Mark McConville, Wilson McLeod & Roibeard Ó Maolalaigh. 2014. *Dlùth is inneach: Linguistic and institutional foundations for Gaelic corpus planning*. Project report. Inverness: Bòrd na Gàidhlig.

- Bermúdez-Otero, Ricardo. 2012. The architecture of grammar and the division of labour in exponence: The state of the art. In Jochen Trommer (ed.), *The phonology and morphology of exponence* (Oxford Studies in Theoretical Linguistics 41), 8–83. Oxford: Oxford University Press.
- Bivand, Roger & Gianfranco Piras. 2015. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software* 63(18). 1–36. <https://www.jstatsoft.org/v63/i18/>. DOI: <https://doi.org/10.18637/jss.v063.i18>
- Borgstrøm, Carl Hjalmar. 1940. *The dialects of the Outer Hebrides* (A linguistic survey of the Gaelic dialects of Scotland 1). Norsk Tidsskrift for Sprogvædenskap, suppl. bind I. Oslo: Norwegian Universities Press.
- Borgstrøm, Carl Hjalmar. 1941. *The dialects of Skye and Ross-shire*. Norsk tidsskrift for sprogvædenskap, suppl. bind II. Oslo: Norwegian Universities Press.
- Bosch, Anna R. K. 2006. Scottish Gaelic dialectology: A preliminary assessment of the Survey of the Gaelic Dialects of Scotland. *Lingua* 116(11). 2012–2022. DOI: <https://doi.org/10.1016/j.lingua.2004.09.001>
- Bosch, Anna R. K. & James M. Scobbie. 2009. Fine-grained morpho-phonological variation in Scottish Gaelic: Evidence from the Linguistic Survey of Scotland. In James N. Stanford & Dennis R. Preston (eds.), *Variation in indigenous minority languages* (IMPACT: Studies in Language and Society 25), 347–368. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/impact.25.18bos>
- Breit, Florian. 2019. *Welsh mutation and strict modularity*. London: University College London dissertation.
- Burnham, Kenneth P. & David R. Anderson. 2004. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research* 33(2). 261–304. DOI: <https://doi.org/10.1177/0049124104268644>
- Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. 2nd edn. (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511805103>
- Cole, Beth. 2015. *Morphosyntactic variation in Uist Gaelic: A case of language shift?* Aberdeen: University of Aberdeen dissertation.
- Cox, Richard A. V. 2017. *Geàrr-ghràmar na Gàidhlig* [A concise grammar of Gaelic]. Ceann Drochaid: Clann Tuirc.
- De Schutter, Georges, Boudewijn van den Berg, Ton Goeman & Thera de Jong. 2005. *Morphological atlas of Dutch dialects*. Vol. 1: *Plural formation of nouns, formation of diminutives, gender nouns, adjectives and possessive pronouns*. Amsterdam: Amsterdam University Press. DOI: <https://doi.org/10.5117/9789053566954>
- Dorian, Nancy C. 1973. Grammatical change in a dying dialect. *Language* 49(2). 413–438. DOI: <https://doi.org/10.2307/412461>
- Dorian, Nancy C. 1977. A hierarchy of morphophonemic decay in Scottish Gaelic language death: The differential failure of lenition. *Word* 28(1–2). 96–109. DOI: <https://doi.org/10.1080/00437956.1977.11435851>
- Dorian, Nancy C. 1978a. *East Sutherland Gaelic*. Dublin: Dublin Institute for Advanced Studies.
- Dorian, Nancy C. 1978b. The fate of morphological complexity in language death: Evidence from East Sutherland Gaelic. *Language* 54(3). 590–609. DOI: <https://doi.org/10.2307/412788>
- Duwe, Kurt C. 2003–2013. *Scottish Gaelic local studies*. http://www.linguae-celticae.org/GLS_english.htm.
- Elsie, Robert W. 1983–1984. Lexicostatistics and its application to Brittonic Celtic. *Studia Celtica* 18/19. 110–127.

- Elsie, Robert W. 1986. *Dialect relationships in Goidelic*. Hamburg: Helmut Buske Verlag.
- Frenda, Alessio S. 2011. Gender in Irish between continuity and change. *Folia Linguistica* 45(2). 283–316. DOI: <https://doi.org/10.1515/flin.2011.012>
- Gillies, William. 1993. Scottish Gaelic. In Martin J. Ball & James Fife (eds.), *The Celtic languages*, 145–227. London & New York: Routledge.
- Gillies, William. 1997. The history of the Survey. In. *Survey of the Gaelic Dialects of Scotland*. Cathair Ó Dochartaigh (ed.). Vol. 1. Dublin: Dublin Institute of Advanced Studies, 25–47.
- Glaser, Elvira. 2013. Area formation in morphosyntax. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szemrecsanyi (eds.), *Space in language and linguistics: Geographical, interactional, and cognitive perspectives*, 196–221. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110312027>
- Goebl, Hans. 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Goebl, Hans. 2010. Dialectometry and quantitative mapping. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation*, Vol. 2. 433–457. New York: Walter de Gruyter. DOI: <https://doi.org/10.1515/9783110219166.1.433>
- Goeman, Ton, Marc van Oostendorp, Piet van Reenen, Oele Koornwinder & Boudeijn van den Berg. 2009. *Morphological atlas of Dutch dialects*. Vol. 2: *Commentary*. Amsterdam: Amsterdam University Press.
- Grieve, Jack. 2014. A comparison of statistical methods for the aggregation of regional linguistic variation. In Benedikt Szemrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*, 53–88. Berlin: de Gruyter.
- Hamp, Eric P. 1989. On signs of health and death. In Nancy C. Dorian (ed.), *Investigating obsolescence: Studies in language contraction and death*, 197–210. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511620997.017>
- Hamp, Eric P. 1997. Background and European setting. In. *Survey of the Gaelic dialects of Scotland*. Cathair Ó Dochartaigh (ed.). Vol. 1. Dublin: Dublin Institute for Advanced Studies, 3–23.
- Hannahs, S. J. 2011. Celtic mutations. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. 5. Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9781444335262.wbctp0117>
- Heeringa, Wilbert & Frans Hinskens. 2014. Convergence between dialect varieties and dialect groups in the Dutch language area. In Benedikt Szemrecsanyi & Bernard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech* (Linguae & Litterae 28), 26–52. Berlin: De Gruyter.
- Heeringa, Wilbert, Martijn Wieling, Boudeijn van den Berg & John Nerbonne. 2009. A quantitative examination of variation in Dutch Low Saxon morphology. In Alexandra N. Lenz, Charlotte Gooskens & Siemon Reker (eds.), *Low Saxon dialects across borders: Niedersächsische Dialekte über Grenzen hinweg* (Zeitschrift für Dialektologie und Linguistik Beihefte 138), 195–216. Stuttgart: Franz Steiner Verlag.
- Husson, François, Sébastien Lê & Jérôme Pagès. 2011. *Exploratory multivariate analysis by example using R*. Boca Raton, LA: CRC Press.
- Iosad, Pavel. 2014. The phonology and morphosyntax of mutation in Breton. *Lingue e linguaggio* 13(1). 23–42. DOI: <https://doi.org/10.1418/76998>
- Iosad, Pavel & William Lamb. 2020. *Morphology and dialectology in the Linguistic Survey of Scotland*. <https://osf.io/4bp2y/>.

- Jackson, Kenneth Hurlstone. 1958. The situation of the Scottish Gaelic language, and the work of the Linguistic Survey of Scotland. *Lochlann* 1. 229–234.
- Jackson, Kenneth Hurlstone. 1967. Palatalisation of labials in the Gaelic languages. In Wolfgang Meid (ed.), *Beiträge zur Indogermanistik und Keltologie: Julius Pokorny zum 80. Geburtstag gewidmet*, 179–192. Innsbruck: Sprachwissenschaftliches Institut der Universität Innsbruck.
- Jackson, Kenneth Hurlstone. 1968. The breaking of original long ē in Scottish Gaelic. In James Carney & David Greene (eds.), *Celtic studies: Essays in honours of Angus Matheson*, 65–71. London: Routledge.
- Kennard, Holly. 2014. The persistence of verb second in negative utterances in Breton. *Journal of Historical Linguistics* 4(1). 1–39. DOI: <https://doi.org/10.1075/jhl.4.1.01ken>
- Kennard, Holly. 2019. Morphosyntactic and morphophonological variation in Breton: A cross-generational perspective. *Journal of French Language Studies* 29(2). 235–263. DOI: <https://doi.org/10.1017/S0959269519000115>
- Kennard, Holly & Aditi Lahiri. 2017. Mutation in Breton verbs: Pertinacity across generations. *Journal of Linguistics* 53(1). 113–145. DOI: <https://doi.org/10.1017/S0022226715000420>
- Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics (EACL'95)*, 60–66. San Francisco: Morgan Kaufmann. DOI: <https://doi.org/10.3115/976973.976983>
- Kortmann, Bernd. 2013. How powerful is geography as an explanatory factor of variation?: Areal features in the Anglophone world. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szemrecsanyi (eds.), *Space in language and linguistics: Geographical, interactional, and cognitive perspectives*, 165–194. Berlin: De Gruyter.
- Lamb, William. 2008. *Scottish Gaelic speech and writing: Register variation in an endangered language* (Belfast Studies in Language, Culture and Politics 16). Belfast: Cló Ollscoil na Banríona.
- Lê, Sébastien, Julie Josse & François Husson. 2008. FactoMineR: A package for multivariate analysis. *Journal of Statistical Software* 25(1). 1–18. DOI: <https://doi.org/10.18637/jss.v025.i01>
- MacInnes, John. 2006. Cainnt is cànan [Speech and language]. In Michael Steven Newton (ed.), *Dùthchas nan Gàidheal: Selected essays of John MacInnes*, 120–130. Edinburgh: Birlinn.
- Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert & Kurt Hornik. 2015. *cluster: Cluster Analysis Basics and Extensions*. Version 2.0.3.
- McManus, Damian. 1994. An Nua-Ghaeilge Chlasaiceach [Classical Modern Irish]. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breathnach (eds.), *Stair na Gaeilge: In ómós do Pádraig Ó Fiannachta* [History of the Irish language: In honour of Pádraig Ó Fiannachta], 335–445. Maigh Nuad: Coláiste Phádraig.
- National Records of Scotland. 2015. *Scotland's Census 2011: Gaelic report (part 1)*.
- Nerbonne, John & Wilbert Heeringa. 2010. Mapping dialect differences. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation*, vol. 1, 550–567. New York: Walter de Gruyter.
- Ó Cuív, Brian. 1983. *The linguistic training of the mediaeval Irish poet*. Dublin: Dublin Institute of Advanced Studies.
- Ó Dochartaigh, Cathair. 1987. *Dialects of Ulster Irish*. Belfast: Institute of Irish Studies, Queen's University of Belfast.
- Ó Dochartaigh, Cathair (ed.). 1994–1997. *Survey of the Gaelic dialects of Scotland*. Dublin: Dublin Institute for Advanced Studies.

- Ó Maolalaigh, Roibeard. 1997. *The historical short vowel phonology of Gaelic*. Edinburgh: The University of Edinburgh dissertation.
- Ó Muircheartaigh, Peadar. 2014. *Gaelic dialects present and past: A study of modern and medieval dialect relationships in the Gaelic languages*. Edinburgh: The University of Edinburgh dissertation.
- Ó Murchú, Máirtín. 1989. *East Perthshire Gaelic: Social history, phonology, texts and lexicon*. Dublin: Dublin Institute for Advanced Studies.
- O'Rahilly, Thomas F. 1936. *Irish dialects: Past and present. With chapters on Scottish and Manx*. Dublin: Browne & Nolan.
- Robertson, Charles M. 1897–1898. Peculiarities of the Gaelic as spoken in Perthshire. *Transactions of the Gaelic Society of Inverness* 22. 4–44.
- Robertson, Charles M. 1898–1899. Skye Gaelic. *Transactions of the Gaelic Society of Inverness* 23. 54–89.
- Robertson, Charles M. 1907. Scottish Gaelic dialects. *Celtic Review* 3. 97–113. DOI: <https://doi.org/10.2307/30069903>
- Scherrer, Yves & Phillip Stoeckle. 2016. A quantitative approach to Swiss German: Dialectometric analyses and comparisons of linguistic levels. *Dialectologia et Geolinguistica* 24(1). 92–125. DOI: <https://doi.org/10.1515/dialect-2016-0006>
- Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de linguistique romane* 35. 335–357.
- Spruit, Marco René. 2008. *Quantitative perspectives on syntactic variation in Dutch dialects*. Amsterdam: University of Amsterdam dissertation.
- Stewart Jr, Thomas W. 2004. *Mutation as morphology: Bases, stems and shapes in Scottish Gaelic*. Columbus: The Ohio State University dissertation.
- Swingle, Kari. 1993. The Irish and other mutations. *WCCFL* 11. 451–466.
- Szmrecsanyi, Benedikt. 2011. Corpus-based dialectometry: A methodological sketch. *Corpora* 6(1). 45–76. DOI: <https://doi.org/10.3366/cor.2011.0004>
- Szmrecsanyi, Benedikt. 2013. *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511763380>
- Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119(11). 1643–1663. DOI: <https://doi.org/10.1016/j.lingua.2007.09.016>
- van der Ham, Margreet, Johan van der Auwera, Sjef Barbiers, Eefje Boef, Hans Bennis & Gunther De Vogelaer. 2005. *Syntactische atlas van de Nederlandse dialecten*. Amsterdam: Amsterdam University Press.
- van der Loo, M. P. J. 2014. The stringdist package for approximate string matching. *The R Journal* 6. 111–122. <https://CRAN.R-project.org/package=stringdist>. DOI: <https://doi.org/10.32614/RJ-2014-011>
- Watson, Seosamh. 2010. Hebridean and mainland dialects. In Moray Watson & Michelle Macleod (eds.), *The Edinburgh companion to the Gaelic language*, 108–127. Edinburgh: Edinburgh University Press.
- Wieling, Martijn & John Nerbonne. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1(1). 243–264. DOI: <https://doi.org/10.1146/annurev-linguist-030514-124930>
- Wieling, Martin, Jelena Prokić & John Nerbonne. 2009. Evaluating the pairwise string alignment of pronunciations. In *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education*, 26–34. Association for Computational Linguistics. DOI: <https://doi.org/10.3115/1642049.1642053>

- Withers, Charles W. J. 1984. *Gaelic in Scotland 1698–1981: The geographical history of a language*. Edinburgh: John Donald.
- Wolk, Christoph & Benedikt Szmrecsanyi. 2016. Top-down and bottom-up advances in corpus-based dialectometry. In Marie-Hélène Côté, Remco Knooihuizen & John Nerbonne (eds.), *The future of dialects: Selected papers from Methods in Dialectology XV*, 225–244. Berlin: Language Science Press. DOI: <https://doi.org/10.17169/langsci.b81.152>
- Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, LA: Chapman & Hall/CRC.

How to cite this article: Iosad, Pavel and William Lamb. 2020. Dialect variation in Scottish Gaelic nominal morphology: A quantitative study. *Glossa: a journal of general linguistics* 5(1): 130.1–31. DOI: <https://doi.org/10.5334/gjgl.1023>

Submitted: 13 June 2019 **Accepted:** 03 August 2020 **Published:** 30 December 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Glossa: a journal of general linguistics* is a peer-reviewed open access journal
published by Ubiquity Press.

OPEN ACCESS 