Transactions of the Philological Society | PhilSoc The Philological Society

Transactions of the Philological Society Volume 00 (2025) 1–36 doi: 10.1111/1467-968x.70011

THE INTEGRATION OF NORSE-DERIVED TERMS IN ENGLISH: EFFECTS OF FORMAL SIMILARITY[1]

By Sara M. Pons-Sanz and Seán Roberts
School of English, Communication and Philosophy, Cardiff University

(Submitted: 27 January, 2025; Accepted: 16 October, 2025)

Abstract

Language change arising from language contact is a complex phenomenon. Peter Matthews encouraged researchers to consider it as firmly grounded in the behaviour of individual speakers. We apply this perspective to investigate the integration of Norse-derived terms into medieval English, testing for the effect of their phonetic similarity to their English (pseudo)cognates. Controlling for etymology, meaning and textual attestation and using a novel data transcription method and coding scheme, we have compiled a database of hundreds of Norse-derived and native forms. The proportion of Norse-derived terms compared to their English (pseudo)cognates is used as a measure of integration; we apply three measures of formal similarity (regarding sound class, articulatory features and historical change). We have found a significant negative relationship: Norse-derived terms are more frequent if they are more similar to their English equivalents. This effect is strongest at the extremes of similarity. There are also significant effects related to the impact of alliteration on the choice of Norse-derived terms and whether forms might have been perceived as more characteristically Norse. Moreover, our results suggest that the rate of use of Norse-derived terms can be taken as an important pointer of the chronological and dialectal origin of a text/manuscript.

1. Introduction

In the monograph *How Languages Can Converge through Contact*, which was still a draft at the time of his untimely death, Peter Matthews made very clear that he was interested in exploring the actual workings of language contact. He was keen to emphasise that one should start with the speakers themselves, even if abstractions need to follow later. For instance, when discussing the introduction of the German loanword *'blitz'* into English, he pointed out that

> [t]he 'donor language' itself, which is another disembodied entity, literally did nothing [...] If English as a language has been subject to an 'influence', it is both over time and through the creative usage of successive groups of its own speakers (§3.3).[2]

This article is submitted as a contribution to the special issue in honour of Peter Matthews, to be edited by Sylvia Adamson and Nigel Vincent.

[1] We are very thankful to Sylvia Adamson, Richard Dance, Add Putter, Nigel Vincent, Alison Wray and the anonymous reviewers for their comments on previous versions of this paper as it has significantly improved thanks to them. Any shortcomings are solely our own.

[2] We are very thankful to Sylvia Adamson and Nigel Vincent for having shared the draft with us.

Accordingly, in this tribute to Matthews's career, we would like to maintain his focus on language users, while bringing together our different approaches to language contact, to provide a novel analysis of the integration of Norse-derived terms into English, using the latest advances in etymology and evolutionary linguistics.

The impact that early medieval Anglo-Scandinavian linguistic contact had on the development of the English lexicon has been the topic of scholarly enquiry since the late nineteenth century,[3] but its study has moved forward significantly since the turn of this century thanks to renewed focus on a number of closely interrelated issues: the etymological identification of the Norse-derived terms in English (Dance 2003; Pons-Sanz 2013; Dance 2019; Dance et al. 2019; Pons-Sanz 2024; see further below, Section 2.2);[4] the closeness between Old English and Old Norse (in terms of linguistic typology and, strongly linked to this, the phonology, morphology and lexis of the two languages), which is likely to have facilitated mutual intelligibility between the speakers of the two languages (e.g. Townend 2002; Keller 2020); and the process by which Norse-derived terms were integrated into English, regarding both their incorporation into particular semantic fields (e.g. Skaffari 2009; Durkin 2014; Pons-Sanz 2015, 2024 and 2025) and their adaptation into English morphosyntactic patterns (e.g. Elter 2024 and forthcoming; Elter & Shaw 2025).

Our main aim in this paper is to further our understanding of the process of integration of the Norse-derived terms into medieval English by analysing the possible impact of formal similarity between these terms and their native equivalents, an area of research that has remained unexplored until now. In the initial stages, when the Norse-derived terms originally made their way into English, the phonological closeness between Old Norse and Old English is also likely to have been a very important factor facilitating linguistic transfer (cf. Dance & Pons-Sanz 2025), particularly given the general trend for loanwords to undergo formal changes in order to fit the phonological structure of the recipient language.[5] However, because of the difficulties involved in accessing relevant data from the period of linguistic transfer (see below, no. (ii)), in this paper, we focus instead on the later use of the terms, once they had become part of the English lexicon. We take a cultural evolutionary approach: in keeping with Matthews's interests, we assume that evolution happens at the level of individual speakers, who 'replicate' a word or not depending on various features of the language production context (e.g. Croft 2000). Our approach is further inspired by recent studies in psycholinguistics which have shown that awareness of the existence of cognates influences how individuals choose lexical items shared by the various languages they are familiar with (e.g. Rogers et al. 2015; Otwinowska 2016; Bosma et al. 2019). Accordingly, our starting hypothesis was that those Norse-derived terms which are formally closer to their native equivalents will exhibit higher levels of use than those terms whose connection with a native term is harder to establish because of their formal differences.

Admittedly, the findings in studies based on the language of contemporary informants cannot be straightforwardly applied to the outcomes of Anglo-Scandinavian contact, but we believe that such comparisons can still be fruitful:

i. As noted above, it is likely that speakers of Old English and Old Norse were able to communicate with each other to an extent without the need to learn the other language.

[3] See Pons-Sanz (2013: 12–14) for an overview of the development of the field; see also Townend (2009, 2024 and 2025) on the role of the wider community of amateur linguists in the early years.

[4] In this paper we use the broader tag *Norse-derived terms* instead of the narrower *Norse loanwords* because we also consider words where Norse influence might lie in the substitution of specific sounds/morphemes rather than the borrowing of the whole term, as well as words that might have been coined in English on the basis of a Norse root (see Section 2.3). On the Norse-derived terms in Scots, originally a form of medieval northern English, see Kries (2003).

[5] For further nuance on this initial process of adaptation, see for instance Peperkamp (2004) and Poplack et al. (2020).

Accordingly, we can easily speak about *societal* bilingualism in the Danelaw (in a nutshell, the areas to the north and east of an imaginary line from London to Chester, where the Scandinavians settled from the ninth century onwards) during the period when both languages were still spoken.[6] Even though there would have been much variation in the level of *individual* bilingualism across time and communities, the significant presence of non-technical Norse-derived terms in Middle English texts is commonly attributed to imposition – or the transfer of vocabulary by Old Norse speakers into English as a result of language shift – and this implies a period of individual bilingualism/second language learning (e.g. Walkden 2025: 126–27; Dance & Pons-Sanz 2025: 231–32, with references).

ii. It is most likely that widespread use of Old Norse in England (as opposed to sporadic encounters with Norse speakers) had ceased by *ca*. 1200 (much earlier in some speech communities than others; see e.g. Parsons 2001; Pons-Sanz 2004). However, it is only during the Middle English period (*ca*. 1150–1500) that we can see the extent of the linguistic impact of this contact in English because of the paucity of earlier sources and the fact that most of them come from areas outside the Danelaw. Accordingly, most of our relevant sources date from a period when we cannot assume the existence of a notable number of Anglo-Scandinavian bilinguals. Therefore, in contrast to studies on living bilinguals, the incidence of Norse-derived terms recorded in our corpus of Middle English texts cannot be attributed to bilingualism *per se*. Nonetheless, we are still dealing with different lexical variants that an individual would have been familiar with and would have associated with different linguistic varieties (dialectal, diaphasic, etc.) rather than different languages.

iii. The reliability of our results is necessarily lower than in studies based on contemporary uses of language: as is generally the case in the investigation of medieval texts, our access to the idiolect of a particular speaker is limited, patchy and – since we focus on scribal practices – constrained by authorial decisions (cf. Benskin & Laing 1981), which have in turn been influenced by stylistic needs (see below, Section 2.1). Because of this and so as to have enough data to carry out statistical tests, we have had to combine the data from the various texts in our corpus to represent the language of speakers from the Scandinavianised areas, though we have also split the sources into different dialectal areas for further investigation (see Sections 2.6 and 3.2). This, in turn, increases the already significant amount of formal variation in the terms under analysis, an issue that is much less relevant in standardised languages.

We discuss these limitations and how we have handled them in Section 2, where we explain our methodological decisions; Section 3 outlines our results in terms of our main hypothesis and other related considerations; Section 4 highlights some conclusions. We argue that our results are generally in keeping with those based on the lexical choices of contemporary bilinguals and that our initial hypothesis is largely upheld, as there is a statistically significant correlation between the formal closeness of Norse-derived terms and their native equivalents, and the proportionate rate of use of the former in English texts.

2. Methodology

To test our hypothesis we have compared the level of use of Norse-derived terms and formally close (near-)synonyms in a representative sample of Middle English texts originating from Scandinavianised areas. We outline below our corpus (Section 2.1), the criteria behind our lexical choices (Section 2.2), our coding practices (Section 2.3), the methods used to calculate

[6] For a recent study on the Danelaw boundary, see Thomas (2019).

the lexical distance between the Norse-derived terms and their native equivalents (Section 2.4) and the statistical methods we have relied on to handle the scores obtained through those calculations (Section 2.5).

2.1. *Corpus*

Our textual choices have been guided by a number of requirements to balance rigour and feasibility: (a) The text records at least some of the terms selected for analysis (see Section 2.2). (b) The text was composed in, and recorded in at least one manuscript from, a (fairly) heavily Scandinavianised area. (c) There is a reliable edition of the version in the relevant manuscript(s). (d) When there are various relevant manuscripts recording the same text, it is easy to differentiate between the versions so as to focus on the choices in a single manuscript.

As the texts have to be searched manually (see Section 2.3), when there was no readily available (suitable) electronic version (e.g. through the *Corpus of Middle English Prose and Verse*, hereafter *CMEPV*), we have created one (marked below by an asterisk in front of the editor's name) by scanning the edited text and turning the scan into a searchable PDF or Word document. When searching for the terms, we have compared the electronic and printed versions so as to minimise possible problems arising from optical character recognition (hereafter OCR).

The lines below present the thirteen sources that make up our corpus, organised by dialectal area.

2.1.1. *East Midlands*

(1) *Final Continuation of the Peterborough Chronicle* (hereafter *FCPC*; ed. *Irvine 2004; approx. 2600 words): it is commonly accepted that this set of annals (1132–1154) was copied as a block into Oxford, Bodleian Library, MS Laud Misc. 636 by a scribe working in Peterborough *ca*. 1154, i.e. in the transition period between Old and Middle English (Irvine 2004: xix and cliii; *Linguistic Atlas of Early Middle English*, hereafter *LAEME*, # 149, petchront.tag).

(2) *Ormulum* (ed. Johannesson & Cooper 2023; approx. 122,000 words):[7] this collection of homilies is generally assumed to have been composed *ca*. 1180 in south Lincolnshire (Bourne); however, Cole and Golding, with Rye (forthcoming), have recently questioned this localisation and have argued instead for a place of composition in east Nottinghamshire (cf. Pons-Sanz forthcoming, and below, note 17).[8] The text is only recorded in Oxford, Bodleian Library, MS Junius 1, which is almost certainly an autograph (*LAEME*, # 301, ormt.tag; Cooper & Dekker forthcoming).

(3) *Havelok the Dane* (hereafter *Havelok*; ed. *Smithers 1987; approx. 18,000 words):[9] this romance is normally dated to the late thirteenth century and said to have been composed in Lincolnshire (possibly Lincoln; see Smithers 1987: lxviv–lxxiii and lxxxix), although its sole extant manuscript, viz., Oxford, Bodleian Library, MS Laud Misc. 108, is assumed to have been penned *ca*. 1300–1325 in west Norfolk (see *LAEME*, # 285, havelokt.tag).

(4) *Genesis and Exodus* (hereafter *Genesis*; ed. *Arngart 1968; approx. 29,000 words): this metrical version of some extracts from the Old Testament was composed in an East Midlands

[7] We are very thankful to Andrew Cooper for having provided us with an electronic version of the text for research purposes.

[8] On the sociolinguistic context of the *Ormulum*, particularly in terms of Anglo-Scandinavian contact, see Golding & Rye (forthcoming).

[9] Manual searches have been helped by the electronic version of the edition by Herzman et al. (1999: 73–185), available through the Middle English Texts Series website.

dialect at some point before the first quarter of the fourteenth century, the *terminus ad quem* suggested by the only manuscript where it is recorded, viz., Cambridge, Corpus Christi College, MS 444. The latter is generally attributed to west Norfolk (see *LAEME*, # 155, genexodt.tag).

(5) Robert Mannyng's *Chronicle* (ed. *Sullens 1996; approx. 210,000 words): Mannyng tells us at the end of the *Chronicle* that he completed it in 1338. The *Chronicle* survives in three manuscripts; we have based our analysis on the forms attested in London, Inner Temple Library, MS Petyt 511.7, which is the only manuscript of Mannyng's texts localised in Lincolnshire. It dates from the second half of the fourteenth century, possibly *ca*. 1375 (Sullens 1983: xviii–xxxv and 1996: 22–51; the *Linguistic Atlas of Late Medieval England*, hereafter *eLALME* 2024, Linguistic Profile, hereafter LP, 38).

(6–8) Local documents from Lincolnshire (approx. 30,000 words), Norfolk (approx. 21,000 words) and Nottinghamshire (approx. 15,000 words): these texts, which are part of the *Corpus of Middle English Local Documents* (2020) and include a wide variety of text types, such as financial accounts, notes of lease, sale and rent, receipts, etc., were composed between 1399 and 1525. Our analysis is based only on the lexical items (i.e. we have not taken into consideration personal or place names).

2.1.2. North-West Midlands

(9) Works of the *Gawain*-poet: *Pearl*, *Cleanness*, *Patience*, *Sir Gawain and the Green Knight* (ed. *Andrew & Waldron 2007; approx. 57,000 words):[10] on the basis of their survival in a single manuscript, viz., London, British Library, MS Cotton Nero A.x, as well as dialectal and stylistic features, these four poems are generally attributed to a single author (here referred to as the *Gawain*-poet), who probably worked *ca*. 1380 in Cheshire (Andrew 1997). The language of the Cotton scribe, who probably worked *ca*. 1400, can also be placed in the North-West Midlands (Putter & Stokes 2007, who provide a more nuanced approach than *eLALME* 2024, LP 26).

(10) *St Erkenwald* (hereafter *Erkenwald*; ed. *Turville-Petre 2024; approx. 3200 words): this poem only survives in London, British Library, MS Harley 2250. This manuscript was written in 1477 but the poem is likely to have been composed in the late fourteenth or the early fifteenth century. Because the text shares stylistic features with those attributed to the *Gawain*-poet and the language of the Harley manuscript can be placed in Cheshire (Burrow 1993; *eLALME* 2024, LP 419; Turville-Petre 2024: 4–5), *Erkenwald* has sometimes also been included amongst the *Gawain*-poet's compositions, although shared authorship is not generally accepted (for contrasting views, see, for instance, Benson 1965 and Turville-Petre 2024: 12–19 for arguments against shared authorship; and Borroff 2006 for arguments in favour).

2.1.3. North

(11) *Cursor Mundi* (hereafter *Cursor*; ed. Morris 1874–1893, accessed through *CMEPV*; approx. 100,000 words): the exact date and place of composition of this versified history of the world from Creation to Doomsday remain unknown, although it is likely to originate from early-fourteenth-century northern England (perhaps somewhere between Durham and the border with Scotland; see e.g. Lamberts 1953: 1–4). *Cursor* survives in nine medieval manuscripts (Horrall 1978: 13–23; Thompson 1998: chap. 1). Our analysis is based on Hand

[10] *CMEPV* does not include an edition of *Patience*, and the other texts by the *Gawain*-poet in the corpus follow different editorial practices.

A's contribution to London, British Library, MS Cotton Vespasian A.iii, one of the earliest and most complete manuscripts recording the original composition (rather than the later 'translation' into a Midlands dialect). Hand A, who was probably working in fourteenth-century Yorkshire, is responsible for ll. 1–16748, 16849–17288, 17317–17852, 18029–20064 and 21173–24968 (*eLALME* 2024, LP 18). Our analysis is based on ll. 1–16700.

(12) *Wars of Alexander* (hereafter *Wars*; ed. Skeat 1886, accessed through *CMEPV*; approx. 56,000 words): this text, which is, in the main, a translation of *Historia Alexandri Magni de preliis*, is recorded in two manuscripts: Oxford, Bodleian Library, MS Ashmole 44 (*ca*. 1450); and Dublin, Trinity College, MS 213 (olim MS D.4.12; fifteenth century). The former provides a *terminus ad quem* for the text's composition, which has been attributed various dates starting *ca*. 1360 (Duggan & Turville-Petre 1989: xlii). The original poet is likely to have come from the West Midlands, probably Lancashire (Duggan & Turville-Petre 1989: xxxvi–xlii), while both manuscripts come from northern England (probably the area around Durham; see *eLALME* 2024, LPs 2487 and 476, respectively). The disparity between the place of composition and copying explains the fact that both manuscripts exhibit a mixture of West Midland and northern forms. As the linguistic features in the version recorded in MS Ashmole 44 are more internally consistent and the text recorded there is longer, this is the version analysed here. Even though Duggan and Turville-Petre's edition is more recent, we have taken Skeat's edition as our starting point because its facing layout makes it easier to differentiate the text in the two manuscripts.

(13) Texts by Richard Rolle (hereafter Rolle; ed. *Hanna 2007; approx. 3000 words): Richard Rolle (*ca*. 1300–1349) was a religious writer from Hampole in south Yorkshire; he composed texts in English and Latin, in verse and prose. The texts that have been surveyed in this study are some of his English prose compositions as copied in Lincoln, Cathedral Library, MS 91 by Robert Thornton, from north Yorkshire. McIntosh (1962: 231) explains that Thornton 'was not by habit a scribe who transformed or "translated" his exemplars so thoroughly as to obliterate all those characteristics in them which were alien to his own' (cf. Hanna 2007: xxxvi–xxxix). Thornton's death *ca*. 1465 provides *a terminus ad quem* for the copying of the texts. While Perry (1866) edited many more prose texts from the Lincoln manuscript which he attributed to Rolle, we have followed instead Hanna's (2007) edition because of its higher reliability in terms of both linguistic matters and authorial attribution. This means, however, that the Rollean compositions explored in this study are much more limited.

Ideally, so as to minimise the impact of metrical constraints on lexical choices, this study would have focused only on texts written in prose because it is widely known that rhyme or alliteration often led an author to choose a form that might have been somewhat peripheral in their active repertoire; and that, similarly, forms in those positions tend to be more resistant to scribal substitutions (cf. Putter 2000: 81; Stenroos 2002; and Putter & Stokes 2007), although they are not completely safe from them (see e.g. Duggan & Turville-Petre 1989: xliii–li on the practices of the scribe who copied MS Ashmole 44 of *Wars*). However, it has not been possible to work with such an ideal corpus because of the limited availability of prose texts that meet the various requirements we have outlined above. Therefore, our corpus also includes poetic texts with different metrical structures: texts associated with the Alliterative Revival (the *Gawain*-poet's texts, *Erkenwald* and *Wars*), and texts whose metrical structure relies on syllabic count, the alternation between stressed and unstressed syllables and, at times, rhyme (the *Ormulum*, *Havelok*, *Genesis*, Mannyng's *Chronicle* and *Cursor*). This methodological decision is likely to have introduced some bias in our data because the recorded forms might not have been those which either the author or the scribe would have naturally chosen, even though – we assume – they would have been familiar with them. We have tried to account for this bias in our analysis, at least as far as alliteration is concerned (see below, Sections 2.5 and 3.1).

2.2. *Terms for comparison*

The texts discussed in Section 2.1 enable us to establish the distribution – across different dates and dialectal areas – of pairs of words that were of Norse vs. native origin, so as to test the hypothesis that Norse-derived words had a higher level of use if they were formally similar to their native equivalents. We pre-selected sixty-five word pairs on the basis of the following criteria, which relate to the key variables under examination:

(1) Likelihood of Norse derivation: the taxonomy for the classification of the (certain/likely/possible) Norse-derived terms in English put forward by 'The *Gersum* Project: Scandinavian Influence on English Vocabulary' (Dance et al. 2019) on the basis of the preliminary work by Dance (2019) provides the starting point for the selection of relevant terms in this study. It offers the most systematic way to analyse these terms on the basis of the reliability of the different types of evidence relevant to etymological decisions. We present here a very brief overview of the four categories into which the words are split, plus other circumstantial information; for more detail, see the aforementioned studies.

> Category A words have watertight phonological (A1), morphological (in terms of inflectional morphology; A2), or both phonological and morphological evidence (A3) for Norse derivation. These terms are marked with an asterisk when there is a direct native cognate.
> Category B includes words whose root is not attested in English before the period of Anglo-Scandinavian linguistic contact in the early Middle Ages, but is attested solely in Old Norse (B1), or in Old Norse and other Germanic languages (B2).
> Category C words have a root attested in English texts composed before the impact of Anglo-Scandinavian linguistic contact became significant but they can be said to exhibit Norse influence on the basis of their derivational morphology (C1), phonological features (which do not offer as definitive evidence as in Category A words; C2), meaning (C3), presence in a compound/phrase (C4) or frequency of attestation (C5).
> Category D brings together words with a very problematic etymology, because it is not possible to settle on a particular interpretation on the basis of the extant evidence (D1) or because that difficulty is compounded by the fact that the form under analysis is itself not clear, which makes etymological discussions even more complex (D2).

The consonants in Categories B and C are doubled when there is some discrepancy amongst scholars regarding the etymology of the term such that not everyone accepts Norse derivation, and tripled when the extant evidence suggests that there is a better explanation for their origin not involving Norse influence. As Category D words are in themselves very problematic, a very low likelihood of Norse origin is marked by the doubling of the consonant.

Various lowercase letters indicate additional circumstantial information that can have a bearing on etymological decisions:

> a = there is at least one continental West Germanic cognate with the same form, sense or usage that is suggestive of Norse derivation.
> b = the attestation of the term in the toponymic record (either during the Middle Ages or later) is strongly associated with the Scandinavianised areas.
> c = the attestation of the term in the lexical record (either during the Middle Ages or later) is strongly associated with the Scandinavianised areas.

d = early occurrences of the term are strongly associated with Scandinavian cultural practices.

In this study, in order to focus solely on reliable cases, we have only considered as possible candidates those terms classified as A, B, or C, without any doubling or tripling of the consonant, in Dance et al. (2019), whose corpus comprises the texts attributed to the *Gawain*-poet, *Erkenwald* and *Wars* (cf. Dance 2019); or Pons-Sanz (2024), where the same principles are applied to the terms that have been suggested to be Norse-derived in the *Ormulum*.

(2) Formal similarity between the Norse-derived term and a native near-synonym: we have focused on pairs comprising cognates (e.g. ME *nout(e* 'ox, bull, cattle' – cf. OIc *naut* – and ME *nēt* < OE *nēat*, from PGmc **nautan*; and ME *aue* 'fear' – cf. OIc *agi* – and ME *eie* < OE *ege*, an *n*- and *i*-stem noun, respectively, based on PGmc **ag-*); or words that might not be direct cognates but are formally very close, here referred to as *pseudocognates* (e.g. ME *thē̆then* 'from that place, thence' – cf. OIc *þaðan*/ONorw. *þeðan* < PGmc **það-* – and ME *thenne* < OE *þanan(e)*, *þanon(e)* < PGmc **þan-*; see Dance 2019: I, 104 and 117).[11] Given that we are interested in measuring the formal distance between the terms in each pair, we have only taken into consideration pairs where both items are attested in our corpus. For example, we have not considered the contrast between forms for ME *gest* with an initial velar (cf. OIc *gestr*) or palatal consonant (< OE WS *giest*, *gyst*, A *gest*) because it is not clear whether the palatal forms are attested in our corpus: in *Genesis*, the only text where they could be present, <g> can represent both [g] and [j] (cf. <gisce> and <gisse> for ME *yī̆tsen* 'to desire') and, accordingly, it is not possible to establish the nature of the initial consonant. It is probably the case that *Genesis* records the velar forms because Middle English texts do not seem to otherwise record palatal forms (see Dance 2019: II, 12; and the *Middle English Dictionary*, hereafter *MED*, 2018: s.v. *gest*, where <yest> is noted as an error) but this argument is somewhat circular. Similarly, our texts only record forms of ME *brinie* 'coat of mail, corselet' without metathesis (cf. OIc *brynja*), while metathesised forms (< OE *byrne*) are recorded, for instance, in Laȝamon's *Brut* (<burne(n)>). While, in principle, we could have calculated formal differences on the basis of those forms, we have avoided doing so for the sake of consistency and, in this particular case, because the predominance of non-metathesised forms might also be attributable, at least partially, to the existence of the Germanic-derived Anglo-Norman noun *brunie*.

(3) Semantic overlap between the Norse-derived term and its native (pseudo)cognate: to minimise as much as possible the impact that semantic factors might have had on lexical choices, we have focused our study on pairs of (near-)synonyms. For example, although ME *ēpen* 'to cry out' (cf. OIc *ōpa*) and ME *wēpen* 'to shed tears, sob, cry' (< OE *wēpan*) are cognates, they are not included here because they do not share the same meaning(s). When the members of a pair overlap semantically or functionally only partially, we have just taken into account the relevant contexts. For example, ME *rōthen* means 'to give advice, advise' (cf. OIc *ráða*) while its native cognate ME *rēden* (< OE WS *rǣdan*, A *rēdan*) has a much broader range of meanings (see *MED* 2018: s.v. *rēden*, v.1); we have only considered those contexts where the native verb has a meaning shared by its Norse-derived cognate. Similarly, in all the texts in the corpus except for *Cursor*, ME *nay* 'no' (cf. OIc *nei*) is only used as an interjection in representations of (direct) speech/thought; accordingly, those are the only contexts that are

[11] The following abbreviations are used in front of cited forms: A = Anglian (a dialect of Old English); ME = Middle English; ODan. = Old Danish; OE = Old English; OIc = Old Icelandic; ONorw. = Old Norwegian; OSwed. = Old Swedish; PDE = Present-Day English; PGmc = Proto-Germanic; VAN = Viking Age Norse; WS = West Saxon (a dialect of Old English). Following common practice in discussions on Anglo-Scandinavian contact, Old Icelandic forms are used to represent the suggested Norse etyma for Norse-derived terms because Old Icelandic is the best attested variety of Old Norse; however, this should not be taken as an indication that these are the exact forms involved in the process of linguistic transfer (hence the use of *cf.* in front of the Old Icelandic terms).

taken into consideration for the use of ME *nō* (< OE *nā*/*nō*). However, in *Cursor* we have also recorded the attestations of ME *nay* and *nō* in contexts where they collocate with ME *certes* 'certainly' (cf. l. 18793, a line which, strictly speaking, is outside our corpus but is also attributable of Hand A of the Vespasian manuscript) or *selcouth* 'peculiar, unusual'.

Notably, we do consider derivatives and compounds including a relevant root, and thus allow a term to be compared not only with its direct equivalent but also with other members of its word-field, so long as the root retains its meaning and either the Norse-derived or the native root could have been used: e.g. in the contrast between the Norse-derived non-metathesised forms representing PGmc **brestan*- (ME *bresten* 'to break, shatter', cf. OIc *bresta*) and the native metathesised forms (< OE *berstan*), we have also counted the occurrences of the root with the prefixes ME *at*-, *out*- or *tō̆*-, regardless of whether both roots appear with each of the prefixes, as this is also an indication of their rate of use.

See Appendix A for the list of our word pairs and all the terms that have been taken into consideration in connection with each pair.

2.3. *Coding of the data*

We have searched the terms associated with our sixty-five word pairs manually in our electronic textual versions; automatic searches are not possible because of spelling variation (and additional problems associated with OCR; see Section 2.1), the existence of homonyms and the need to ensure a relevant attestation from a semantic perspective (see no. (3) in Section 2.2). To make the statistical analysis of the *ca*. 9000 tokens retrieved through the searches possible, we have created an Excel spreadsheet where the tokens are coded and classified in connection with the following categories:

(1) Set identifier: this tag marks which forms should be considered together and therefore allows us to compare the attestations for each pair of Norse-derived and native terms. In each case, we have chosen the Norse-derived term as the identifier: e.g. '**nay**' is used to indicate that attestations of Norse-derived ME *nay* should be compared with those of its cognate ME *nō*. In the database we have separated the tokens of strong verbs into different ablaut forms (e.g. infinitive/present forms, (singular/plural) past forms, past participle forms) to capture the distribution of forms across the paradigm in case this is useful for future research. However, the results presented below do not take those differences into account and are based instead on whole-paradigm comparisons for all verbs except for the forms brought together under '**hilen**' (ME *hilen* – cf. OIc *hylja* – vs. ME *hēlen* < OE *helan*), which are separated between infinitive/present, past and past participle forms. The reason for this is that, in the case of ME *bresten*, *gē̆ten* and the velar forms in *yēven* (see below, no. (2)), Norse influence might refer simply to the substitution of specific sounds for the native equivalents rather than borrowing *per se* (cf. Dance 2019: I, 98–99; Dance et al. 2019: s.v. *gif*) and vocalic variation is not directly relevant to the distinction between Norse and native forms because the forms simply exhibit the variation that one would expect in the native strong forms (see below, no. (7)). However, as far as the forms discussed under '**hilen**' are concerned, the distinction between Norse-derived and native forms relies not only on vocalic alternation but also on the fact that OE *helan* was a strong class IV verb and its Middle English reflex exhibits both strong and weak forms, while ME *hilen* is only conjugated with weak forms.

(2) Lexeme: here we follow *MED* (2018), even though it does not always separate Norse-derived and native forms under different headwords: e.g. Norse-derived forms with an initial velar consonant (cf. OIc *gefa*) and native forms with a palatal consonant (< OE WS *giefan*, *gyfan*; A *gefan*, *gifan*) are presented together under ME *yēven* 'to give'. In those cases where the *MED* headword does not clearly represent the Norse-derived or the native form, *cf.* precedes the headword both in our spreadsheet and in Appendix A below.

(3) Word class: the term's word class is recorded so that this can be taken into account when calculating rates of use because word classes are said to occupy different places in scales of borrowability (see further no. (1) in Section 2.6).

(4) Etymology: the tokens are classified as either '**Norse**' or '**English**'. In those cases where the etymology is not clear because of the spelling practices of a text, a question mark is used and the forms are not counted: e.g. <halfsuster> in *Gawain* l. 2464 could be a Norse-derived form (cf. OIc *systir* 'sister') or a native form (ME *suster* < OE *swustor*, variant of *sweostor*) because <u> in MS Nero A.x can occasionally represent [y] instead of [ʊ] (see Dance et al. 2019: s.v. *half-suster*, where this form is classified as CC2).

(5) Recorded spelling: we have collapsed together forms where spelling differences are unlikely to represent phonetic differences (e.g. <ai> and <ay>).

(6) Distinctive sounds: here and in (7) we have phonetically transcribed the recorded spellings manually. In his spelling practices, Orrm distinguishes between two different allophones of /ɡ/: the voiced velar stop [ɡ], which he represents with a flat-topped *g* (we have transcribed it as <ḡ > in the database and below); and the voiced velar fricative [ɣ], which he represents with a composite graph including an Insular *g* and *h* (we have transcribed it as <$\overset{\text{h}}{\text{ʒ}}$>). This suggests that he saw these two sounds as distinctive enough. So as to capture this and to be consistent in our measurements, we have tried to differentiate between all the relevant allophones (including those cases where the phoneme eventually split into two different phonemes because the allophones were no longer in complementary distribution because of various factors):

/f/: [f]/[v] (in voiced environments).
/ɡ/: [ɡ]/[ɣ] (in intervocalic position).
/h/: [h] (initial position)/[x] (after back vowels)/[ç] (after front vowels).
/n/: [n]/[ŋ] (followed by a velar consonant).
/s/: [s]//z/ (in voiced environments).
/θ/: [θ]/[ð] (in voiced environments).

In those contexts where it is not clear which sound(s) the attested spellings represent, we have surrounded each option with round brackets and have separated them with a slanted bracket so as to allow the program to calculate the phonological distance between the forms in connection with all the different options (see below, and Sections 2.4 and 2.5). For example,

(ɛː)/(eː): there are a number of terms where the different reflexes of PGmc */eː/[1] in Old Norse and English provide the main evidence for Norse derivation. In Old Norse this sound evolved as /ɑː/, represented in Norse-derived terms as /ɔː/ (particularly in the South and Midlands) or /aː/ (mainly in the North): e.g. ME *hōr* 'hair' (cf. OIc *hár*); and ME *rōth* 'advise, counsel' (cf. OIc *ráð*). In Old English, however, it evolved as /æː/ in West Saxon and /eː/ in Anglian and Kentish; the reflexes of these two sounds in Middle English are /ɛː/ and /eː/, respectively, both commonly represented by <e>: e.g. ME *hēr* (< OE *hǣr*) and *rēd* (< OE *rǣd*; see Dance 2019: ɪ, 88, with references). Given that our texts represent varieties associated with Old English Anglian dialects, we would expect /eː/ in these terms; however, Orrm's <æ> (<hær> and <ræd>) represents /ɛː/ instead, which points towards the existence of phonological variation and, hence, uncertainty in the transcription.

(ɛɪ)/(aɪ): even though it is commonly suggested that the lowering of [ɛɪ] and its merger with [aɪ] had happened by 1300 (e.g. Jordan 1974: §195), Wełna (2007: 8–10) argues that spellings in *Havelok* and *Genesis* do not evidence the completion of the process. Accordingly, the two alternatives are considered for terms with <ei, ey> in these two texts, while the merger is assumed to be absent from *FCPC* and *Ormulum* (as suggested by

Orrm's consistent use of <eȝȝ> in relevant contexts), and fully completed in the other texts on the basis of the date of the manuscripts.

(ɛw)/(eʊ): late Old English/early Middle English /æ:/ resulting from the monophthongisation of OE /æ:a/ (<ēa>) was common in both West Saxon and Anglian (Hogg 2011: §§5.210 and 5.212), and was raised to /ɛ:/.[12] Jordan (1974: §§107 and 208) explains that, when it appeared next to /w/, this combination eventually gave rise to /ɛʊ/, which often merged with /eʊ/. Accordingly, <e> in <few, fewe> for ME *feue* 'few' (< OE *fēawa*) could represent /ɛ:/ (in disyllabic forms), and [ɛ] or [e] in monosyllabic forms (see below).

(vowel)/(long vowel): we have generally transcribed vocalic length on the basis of the form of the headword in *MED*; in those cases where *MED* marks the vowel as both long and short (e.g. in contexts prone to lengthening because of the presence of a homorganic consonant cluster: e.g. ME *bō̆nd* – cf. OIc *band* – vs. ME *bē̆nd(e* < OE *bend*; see Jordan 1974: §22; Pons-Sanz 2024: 124–27), we have taken both possibilities into account. When *MED* marks the vowel as long on the basis of Middle English Open Syllable Lengthening (Jordan 1974: §§25–26), we have also considered both options, except for the *Ormulum*, where the process has not taken place yet (Cooper & Åberg forthcoming; Goering forthcoming). Similarly, we have accounted for long and short vowels for the Norse-derived (cf. OIc *þó* < VAN *þóh*) and native forms (< OE *þēah*) which *MED* lists under ME *though* because, even though we might have expected long vowels, Orrm's spellings (<þohh> and <þehh>, respectively) point towards short vowels.

The use of various alternatives is associated with the difficulty in establishing whether a historical, unstressed vowel (commonly [ə]) in final or medial position was pronounced or not:[13]

(Vowel +[j] or [w])/(diphthong with [ɪ] or [ʊ] as the second element): our underlying assumption is that, in disyllabic words with the structure (C)VxV, the intervocalic sound marked here with <x> provides the onset of the second syllable and, therefore, we can identify it as a semivowel rather than as the second element of a diphthong; the latter is the structure that we would find in monosyllabic forms. In some cases, we can be certain that we are dealing with a disyllabic or a monosyllabic form. For instance, the regularity of Orrm's metre suggests that for him <eȝȝe> (ME *eie* 'fear, dread' < OE *ege*) is a disyllabic term ([ɛjə]); although the final vowel is elided to avoid vocalic hiatus in the phrase <eȝȝe off dæþ> 'fear of death' (ll. 19962, 19966; see Cooper & Åberg forthcoming), we can assume that [j] is retained when the noun is pronounced together with the preposition. In contrast, <au> for ME *aue* (cf. OIc *agi*) in *Cursor* suggests that we are dealing with a monosyllabic term that exhibits the outcome of the labialisation and vocalisation of the original [ɣ] (Jordan 1974: §186). However, in other cases it is harder to know whether we are dealing with a disyllabic or a monosyllabic form.

[ɣ]/(x) and (ðə)/(θ): these pairs capture the uncertainty regarding the presence or absence of schwa (with subsequent devoicing of the original consonant; see Jordan 1974: §§187, 197 and 203) when <e> appears in the spelling of terms like ME *aue* (when the middle sound is presented with a consonant, e.g. <ȝ, gh>, rather than a vowel) and ME *bōth(e* (cf. OIc *báðir* and OE *bā þā*).

12 On the evolution of OE /æ:/ from different sources in Middle English, see further Stenbrenden (2019).

13 See, for instance, Jordan (1974: §§138–41 and 152–55) and Fulk (2012: 48–50) for an overview; Minkova (1992: 171–73) on the situation in *Genesis*; and Putter et al. (2007): chap. 2) for a discussion of the loss of final schwa in the alliterative texts.

(7) Full form: this is the phonological structure that we have taken into account to calculate the formal distance between the Norse-derived terms and their native equivalents. We have transcribed here the whole root, and have only considered inflexional or derivational affixes when they are part of the distinctive sounds recorded in no. (6): e.g. the past and past participle inflectional suffixes of ME *hilen* and *hēlen* (see no. (1)), and the two different derivational suffixes in ME *eightend* 'eighth' (cf. OIc *áttandi*) and *eightethe* (< OE *eahteoða*). So as not to complicate matters unnecessarily, we have used archiphonemes when there is or there could be variation in the sounds that we are dealing with but this variation affects both the Norse-derived and the native terms and, therefore, it is not important for our purposes:

> [V] = a vocalic sound: e.g. our corpus records a large variety of forms associated with the adverb represented in Present-Day English by *again* and the preposition *against* (both discussed under the set '**again**'). We can see spellings with <æ, e, ai, ay, ey> in Norse-derived and native forms alike. Variation in the native forms reflects the fact there was important dialectal and chronological variation in Old English, as summarised by *MED* (2018: s.v. *onyēn*, prep.) for the prepositional forms: 'Merc. [Mercian] *ongegn*, WS *ongeagn*, *ongēan*, Nhb. [Northumbrian] *ongægn*, *ongæn*; also LOE (prob. in all dialects) *ongēn*'. In the process of adaptation into English, the Norse forms probably had a diphthong (cf. OIc *(í) gegn*; see Dance et al. 2019: s.v. *agayn*, adv., prep.) but the root vowel is likely to have exhibited very soon the same variation as in native forms (e.g. Norse-derived <onngæn> alongside native <onnȝæn> in the *Ormulum*).
>
> [F] = [f]/[v]: loss of [ə] is likely to have caused the devoicing of final [v] in forms of ME <arf> (ME *erve* 'inheritance; cattle') and <ge/if/v-> (cf. ME *yēven* 'to give'; Jordan 1974: §217).
>
> [Θ] = [θ]/[ð]: given the tendency towards the voicing of initial [θ], particularly – albeit not only – in unstressed words (Jordan 1974: §207), the two consonants might have alternated in initial position in pairs such as ME *thē̆then* (obsolete in Present-Day English) vs. ME *thenne* (currently pronounced with [ð]), or ME *thwert* (cf. OIc *þvert*) vs. ME *thurgh* (< OE *þurh*), whose reflexes are both currently pronounced with [θ]. [Θ] is also used to represent the possible alternation of the two sounds caused by the loss of final [ə] (e.g. ME *scāth(e*, cf. OIc *skaða*, vs. ME *shāthe* < OE *sceaþa*).[14]
>
> /W/ = see below: as Laing & Lass (2019) describe, there is much variation in the spelling (our dataset records <wh->, <w-> and <qu->) and the pronunciation of the reflex of PGmc */xw-/ during the Middle English period. They (2019: 94 and 102) propose [xw] > [hw] > [^{h}w] > [w] as the most common evolution, and [kw] > [xw] > [hw] > [w] as the common development for spellings with <q>. In the case of ME *whē̆then* 'from wherever, whence' (cf. OIc *hvaðan*) vs. ME *whenne* (< OE *hwanon*), this variation is likely to have been compounded by the fact that the former is Norse-derived, and it is difficult to know the phonetic nature of the initial consonant at the point of entry into English and its exact evolution.

A full statistical report is available in the supplementary Data S1, and our full dataset and code can be found here: https://github.com/seannyD/IntegrationOfNorseTerms.

2.4. Measuring lexical distance

Lexical distance is a method of measuring the similarity of two wordforms. A simple approach is the 'edit distance', which essentially counts the minimum number of edits to the first form that would need to be made to convert it into the second form (e.g. the Levenshtein

[14] On the relationship between [θ] and [◇] in Middle English, see further León Prol (2025).

distance; see e.g. Wichmann et al. 2010; Heeringa et al. 2023). One way of calculating this is to align the segments of each form to minimise the number of mismatches. For example, the Norse-derived form [axtand] (ME *eightend*) and English [axtθ] (ME *eightethe*) might be aligned like this:

a x t a n d
a x t – – θ

In a simple edit distance calculation, the first three segments are the same, so contribute distances of zero. The last segments differ, so contribute a distance of 1. The middle [a] and [n] in the Norse-derived form have no clear partner in the English form, so are aligned with a gap and contribute a special gap cost (e.g. a distance of 1). The distances between each aligned segment are then added and normalised (e.g. by dividing by the maximum possible score for an alignment of the same length) to give a normalised distance score between 0.0 (two forms are identical) and 1.0 (forms are maximally different). In the example above, the total distance is 3 (the Norse-derived form is '3 edits' away from the native form), and the normalised distance is 0.5 (=3/6).

Our aim is to measure the lexical distance between Norse-derived and native forms in each set on the basis of the full form (no. (7) in Section 2.3) – not just the distinctive features (no. (6) in Section 2.3) – in terms of an optimal alignment framework. Recent studies use specific distances depending on the types of the segments being compared (to model assumptions such as /d/ and /θ/ are 'closer' than /d/ and /k/), and this is the approach we have taken here. To handle this, we use the Needleman-Wunsch algorithm (Needleman & Wunsch 1970), implemented in the python program *lingpy* (List & Forkel 2024). This algorithm was developed to deal with the alignment of genetic sequences and has a similar function to the Levenshtein algorithm, but it makes it possible to specify arbitrary distances between specific pairs of segments. Given these distances, the Needleman-Wunsch algorithm is guaranteed to find an alignment that minimises the total distance.

There are three general approaches to quantifying distances between segments; they are based on sound class, articulatory features and historical change. For the sake of rigour, we have used all three. The sound class approach assigns a high distance when comparing two segments from different sound classes (e.g. a consonant and vowel). In some studies, this is an arbitrary high value (see Nerbonne & Heeringa 2010; Gooskens & Swarte 2017; Gooskens & Van Heuten 2020), while others have a more complex approach. For example, in his study on the mutual intelligibility between speakers of Old Norse and Old English on the basis of the terms included in the Leipzig-Jakarta list (Tadmor et al. 2010), Keller (2020) assigned a distance of 0.5 for alignment between two different vowels or between two different consonants (with an additional distance of 0.25 for differences in vocalic length), but a distance of 1.0 for an alignment between a vowel and a consonant (as well as for insertions and deletions). We adopt Keller's method with one modification to handle cases where the features of a segment are uncertain: in those cases we have assumed that the segment could match any of its possible realisations and have identified the closest alignment of any of those realisations with the other form. For example, when considering the distance between the present roots [hɪl] and [h(ɛ)/(ɛː)l] (ME *hilen* vs. *hēlen*; see above, nos. (1) and (6) in Section 2.3), two comparisons are made: [hɪl] vs. [hɛl] and [hɪl] vs. [hɛːl]. The distance scores for each comparison are calculated, and, as the first comparison has a lower distance (because it does not involve an additional cost from a difference in vocalic length), that distance is chosen to represent the distance between these two forms.

The articulatory feature-based system uses more motivated distances. While sound class methods provide reasonable results, their scoring system is somewhat arbitrary, and they treat all segments within a class as equally similar. For short words, this can create misleading results: for example, using Keller's reckoning, the single-syllable forms [aɪ] (ME *ai* 'always, ever'; cf. OIc *ei*) and [a:] (ME *ō* < OE *ā*) have a normalised distance of 0.75 (0.5 for different vowels + 0.25 for different length), while forms such as [pa] and [bɪ], which are intuitively more distant, have a normalised distance score of 0.5: [p] and [b] are both consonants (score 0.5), and [a] and [ɪ] are both vowels (score 0.5). In contrast, the feature-based scoring system compares segments using articulatory features (see e.g. Ahmed et al. 2021). In our study, each phonetic segment is represented by a set of categorical features as specified by the 'Broad IPA' transcription system made available through the Cross-Linguistic Transcription Systems database (List et al. 2024). Consonants are represented by features of phonation (voiced, voiceless), place (bilabial, labio-dental, dental, alveolar, post-alveolar, palatal, labio-velar, velar), airstream (normal, sibilant, lateral) and manner (affricate, approximant, fricative, trill, nasal, stop). Vowels are represented by features of height (open, close, mid, open-mid, close-mid), centrality (front, central, back), duration (long, short) and roundedness (unrounded, rounded). The distance between two consonants is the Hamming distance between their features (i.e. the proportion of features whose categories differ between the two consonants; Hamming 1950). The distance between two vowels is the Hamming distance between their features or, for diphthongs, the average of the Hamming distances between the start and end points. Segments from different sound classes (e.g. consonants, vowels) have been assigned the maximum distance. For the archiphonemes (see no. (7) in Section 2.3), we have assumed that they matched any of the possible values: e.g. if a segment is transcribed as [Θ], i.e. it could represent either [θ] or [ð] (with uncertainty about voicing), its phonation feature has been set as 'unspecified' and it matches both voiced and voiceless features. Otherwise, as outlined above, ambiguity in transcription has been dealt with by finding the minimum distance between all possible realisations of the form.

Finally, we have also calculated distances between forms on the basis of likelihood of historical change, taking Jäger's (2015) work as our starting point. We have estimated distances between segments by aligning forms from the Automatic Similarity Judgement Program (hereafter ASJP; Wichmann et al. 2022) within cognate sets and using phylogenetic trees to establish the likelihood of each segment being replaced by another. For example, [b] is much more likely to change to [m] than to [d], despite [b] and [d] also sharing many articulatory features. Since they are based on the probability of change, the distances are continuous. Notably, while conceptually appropriate given our corpus, the ASJP uses a coarse set of phoneme categories to represent its raw data and groups together some segments (e.g. voiceless and voiced dental fricatives). In spite of this limitation, it is a useful alternative measure that allows us to estimate the robustness of the results through triangulation. As with our feature-based scores, the archiphonemes have been treated as matching any of the possible realisations; for diphthongs we have taken the average of the distances for each component; and for distances between forms with ambiguous realisations we have taken the minimum distance for all potential forms.

The sound class method applied in Keller (2020) specifies a gap penalty of 1 point of distance. For the other two measures, it is not clear whether this is appropriate. Intuitively, the gap penalty should be proportionate to the typical cost of unmatching pairs (i.e. it should be acceptable to introduce gaps if the rest of the alignment improves as a result), and this is more complex for the calculations based on historical distance. Accordingly, we have used a two-step process: first, we have calculated m, that is, the mean distance between forms in our data with a gap penalty of 1.0. We have then set the final gap penalty as $\alpha \times m$. To choose α, we have experimented with a range of values on some form pairs with ambiguous alignment and have found $\alpha = 0.1$ to produce sensible results. Finally, since the historical distance scores are on an arbitrary scale, the distances have been adjusted so that each measure ranges between 0 (identical forms) and 1 (the largest distance in our sample).

While highly correlated, the three approaches make different assumptions about the distances between segments, which can lead to different alignments. For example, the cognate forms [graɪθ] (ME *greith* 'ready', cf. OIc *greiðr*) and [rɛ:d(ɪ)/(i:)] (ME *rēdī* < OE *gerǣde* + ME -*ĭ̄*) are aligned differently according to the three schemes. The sound class model finds low similarities between most of the segments, but the gap cost is high enough that it avoids adding any gaps, producing the following alignment (scaled distance of 0.648):

g r a ɪ θ
r ɛ: d (ɪ)/(i:)

This is unlikely to be the alignment that a historical linguist would produce; it simply is the alignment that produces the lowest distance given the assumptions of the sound class model. In contrast, the feature-based system prefers to align segments with similar features, so it recognises that [r] and [d] have at least similar places of articulation and aligns the diphthong with the vowel it shares an endpoint with:

g – r a ɪ θ
r ɛ: d (ɪ)/(i:) –

This produces a scaled distance of 0.721, which is slightly higher than the sound class system, though the scaled distances are relative to the rest of the distribution, and the relative rank of the distance in relation to other comparisons in the sample is similar to the sound class result. Finally, the historical system creates an alignment that reflects some of the historical sound correspondences. Accordingly, it produces a scaled distance of 0.705, which is slightly lower than the feature-based distance:

g r a ɪ θ –
– r ɛ: d (ɪ)/(i:)

In general, the historical system deals more parsimoniously with metathesis than the sound class model. For example, for the comparison between [hʊndrVθ] and [hʊndVrd] (cf. ME *hundred*), the sound class model prefers to open no gaps, comparing the final vowel with a consonant:

h ʊ n d r V θ
h ʊ n d V r d

In comparison, the historical system aligns the [r] and [d] correctly, and opens gaps, essentially representing a transposition of the vowel:

h ʊ n d – r V θ
h ʊ n d V r – d

In summary, the three measures of distance use different principles and assumptions. In the analysis below, we analyse all three and show that they all show the same pattern, demonstrating that the conclusions are robust. We have applied the three distance measures as follows. Each set (see no. (1) in Section 2.3) consists of several attested forms of Norse or English origin, and each form has a frequency of occurrence within each source. Therefore, within a set, every pair of Norse-derived form and English form has been compared, and the length-normalised distance has been calculated and entered alongside the frequency of each form as a single observation. That is, a set that includes two Norse-derived forms and two English forms would yield four comparisons. While it would be possible to merge results within a set to a single observation, it is not clear that the mean lexical distance between several pairs is representative of the true distance (e.g. if one comparison is very distant and the other is very similar, the mean would be middling, which would represent neither).

Similarly, it is not clear whether it would be better to combine raw or relative frequencies. Instead, we have considered all pairs of forms to keep the statistical analysis at the level at which the data have been collected, and we have controlled for non-independencies within a set using various statistical methods, outlined in the next section.

2.5. *Statistical methods*

We have used mixed effects modelling in a regression framework to model the data. Regressions try to find a model that best numerically predicts a dependent variable of interest according to a set of independent variables. Statistical modelling has been implemented in R (R Core Team 2021) with the *lme4* package (Bates et al. 2015), using a mixed effects modelling framework that allows us to control for various aspects. We have treated the process of writing manuscripts as a binomial process: when wanting to express a meaning captured by a particular set, the scribe faced a binary choice of using a Norse-derived form or an English form; this choice was perhaps unconscious, but it was also constrained by the presence of a particular form in the scribe's exemplar (cf. no. (iii) in Section 1). The dependent variable was the likelihood of observing the Norse-derived form compared to both Norse and English forms and for each pair in a set. The binomial model accepts the raw frequencies of the Norse and English forms, allowing it to consider both the ratio and the absolute frequency. The main independent variable is the normalised distance measure between the Norse-derived and English forms in the corpus (see Table 1 for an example of some of the data that went into modelling the historical distance). The model includes a random intercept for each set, which controls for the possible influence of individual sets in our sample. Furthermore, because there are non-independencies between the pairs of words compared in a set, we have included a random intercept for each English lexeme.

To test for non-linear relationships in the data, we have made two additions. Firstly, we have created more complex regressions by adding quadratic and cubic transformations of the distance measure. This allows the model to represent single-curved (e.g. rapidly increasing) and double-curved (i.e. s-curved) relationships, respectively, between frequency and form distance. We have used model comparison to test whether these higher-order variables provide a significantly better fit to the data (see e.g. Bates et al. 2015). Secondly, we have used a General Additive Model (hereafter GAM) using the R package *mgcv* (Wood 2011) with the same structure as the mixed effects model. GAMs use non-linear splines to fit the data, providing the model with the freedom to fit arbitrary smooth relationships that are justified by the data, rather than assuming the particular shape of quadratic or cubic curves (see e.g. Monaghan & Roberts 2019; Monaghan & Roberts 2021). The Effective Degrees of Freedom measure of a fitted GAM model indicates whether a fit is close to linear (near 1.0) or more non-linear (higher than 1.0). Since we do not have a theoretical reason for very high non-linearity, the number of spline knots has been limited to 4. The linear mixed effects framework and the GAM framework have slightly different assumptions, allowing us to test the robustness of the findings.

Table 1. Example of some of the data used in the statistical analysis

Set	Norse form	English form	Normalised historical distance	Norse frequency	English frequency
blest	blast	blɛst	0.060	13	1
bond	b(a)/(aː)nd	b(ɛ)/(ɛː)nd	0.107	44	2
bond	b(ɔ)/(ɔː)nd	b(ɛ)/(ɛː)nd	0.141	16	2
both	bɔːðən	bɔː	0.994	18	13

The method above has been applied to two perspectives on the data. The first analysis combines the total frequency across all sources, which provides the most reliable estimates of frequency balance between Norse-derived and English forms. The second analysis examines frequency observations at the source level, meaning that details about the source can be taken into consideration as additional controls including a random intercept for the source. Since the proportion of forms may change over time, we have included an independent variable representing the age of the source. As it is often not possible to establish the exact date of many of our manuscripts, we have simply used the century when they were written as an ordinal variable; given their date ranges, we have associated the documents from Lincolnshire, Norfolk and Nottinghamshire with the fifteenth century. Since some of the sources are alliterative poems, lexical choices are likely to have been affected by metrical requirements; after all, alliteration is often noted as one of the factors that are likely to have contributed to the lexical richness of these texts (see e.g. Dance et al. 2025: 32). The influence of an alliterative metre would only affect pairs of forms that did not alliterate (forms that alliterate are not in competition). Accordingly, we have written an algorithm to identify whether the source is an alliterative poem, and whether or not the two forms alliterate (considering consonant clusters, as well as alliteration between vowels and [h]); we have added this variable as a binary fixed effect in the source-level regression.

2.6. *Exploratory analyses*

In addition to the hypothesis testing methods discussed in Section 2.5, we have also conducted three exploratory analyses:

(1) Since historical integration appears to be more complex for some word classes (e.g. verbs vs. nouns or adjectives, or grammatical words vs. lexical words; cf. e.g. Winford 2003: 51; Tadmor et al. 2010: 231), we have tested whether word class predicts the frequency of use of Norse-derived terms in our corpus. For each of the nine word classes, a binary variable indicates whether the given forms could be used as that class (some forms could be used as multiple classes: e.g. the forms that we have included in the set '**again**' can be used as an adverb or a preposition). Following previous work, we have made two additions to the total frequency regression model above (in addition to form distance and controls for set and lexeme): adding only the binary variable for verbs, and adding each of the nine word class variables. If considering word class significantly improves the model's predictions, this would suggest that integration might work differently for different word classes.

(2) The distance measures explained above take into account any differences in form, but some differences could have been perceived as more distinctive than others because of their prevalence in Norse-derived terms vs. the general evolution of native terms: e.g. lack of palatalisation of [k], [g] and [sk] in contrast with the absence of metathesis of /r/ + vowel. After all, in Old English there is already some alternation between forms with and without metathesis (see Hogg 2011: §§7.94–95).[15] We have singled out the sets of forms whose diagnostic contrast involves the presence of a root-initial velar vs. a palatalised consonant, with specific focus on [sk] vs. [ʃ] because the palatalisation of that cluster in native terms took

[15] On the absence of palatalisation of velar consonants and the absence of metathesis as diagnostic of Norse-derivation, see Dance (2019: I, 95–100 and 108, with references). Given the dates of our texts, and as noted above (no. (ii) in Section 1), distinctiveness refers not so much to the speakers' awareness of the terms' Norse origin but to their awareness of the existence of sets of alternative forms that might have been perceived as being marked in terms of dialectal (and perhaps other types of sociolectal) distribution. The distinctiveness of these features can be thought of in terms of degrees of congruity. In the case of the distinction between root-initial velar vs. palatal sounds, it is also important to bear in mind that some studies have shown that listeners rely more on word beginnings than on later parts for word recognition, although results are not consistent. See further Gooskens (2024: 125 and 128, with references).

place beyond the common limits of the process (e.g. even in front of back vowels in initial position; Hogg 2011: §7.16). We have entered this variable (which we call the *Norse Diagnostic Score* in Section 3.2) as a fixed effect into the regression explained above.

(3) We have also evaluated differences between manuscripts in terms of the proportion of English and Norse-derived forms in each set to explore how it correlates with dialectal and diachronic variation. After all, our data offer more detail in this respect than the linguistic profiles in *eLALME* or discussions about the dialectal origin of a text/manuscript, where the presence of (particular) Norse-derived terms is commonly presented as one of the indicators of origin in a (heavily) Scandinavianised area (cf. Duggan & Turville-Petre 1989: xxxvi–xxxvii and xli). For each manuscript, we have calculated the proportion of Norse-derived forms compared to all forms (P_N) for each set. The mean P_N for all sets (MP_N) has been used as a measure of Norse proportions for that manuscript. Then, for every possible pair of manuscripts, the difference between MP_N for the two manuscripts has been taken as a measure of the distance between the two manuscripts. To visualise the results, we rely on a neighbour net using the R package *phangorn* (Schliep 2011; see Section 3.2). This is a method borrowed from genetics that shows distances between pairs of items in a network (cf. McMahon & McMahon 2005).

3. Discussion of results

3.1. *Effect of formal similarity*

The final data include 537 attested forms in sixty-five lexical sets across thirteen sources, which produces 1636 comparison pairs of Norse-derived and English forms. For the total frequency data, the proportion of Norse-derived forms for each comparison ranges between 0% and 100% (mean = 56.6%, SD = 31.3%). Figure 1 shows the distribution of frequencies in our corpus. Notably, even though the *Ormulum*, one of our earliest sources, is frequently singled out as a text exhibiting heavy Norse influence, when it comes to the rates of attestation of Norse-derived terms with close formal similarity to English (near-)synonyms, its scores are

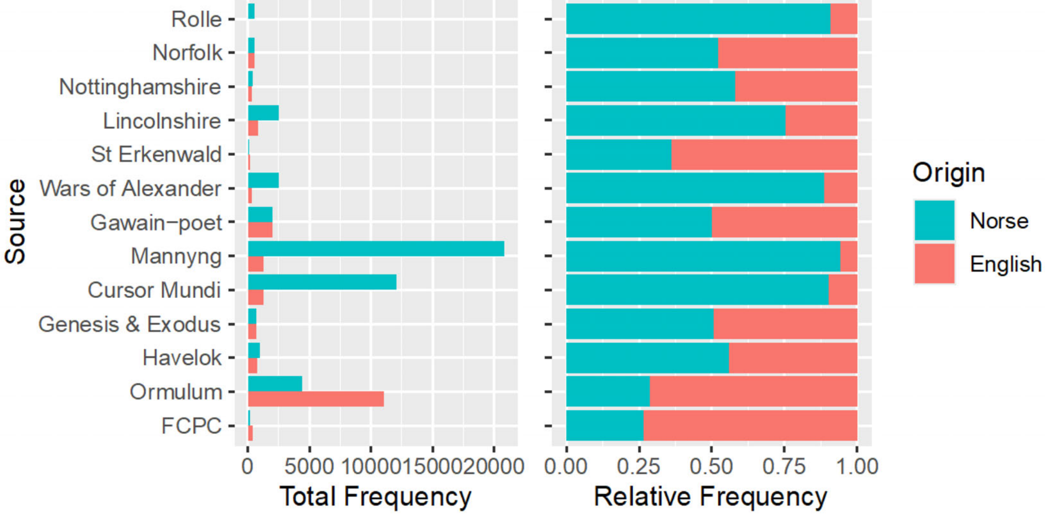

Figure 1. The total frequency (left) and relative frequency (right) for Norse-derived and English forms from each source in the corpus.

lower than later compositions, possibly because in the later texts Norse-derived terms had longer to become embedded and gain ground on native forms. As noted in Pons-Sanz (2024: chap. 4), the distribution would probably look very different if we had compared the rate of use of Norse-derived terms in relation to *all* their (near-)synonyms because French-derived terms would have been much more prevalent in later compositions.

Figure 2 shows some descriptive statistics for the form distance measures, making clear that the three approaches used to calculate them are highly correlated with each other. For the sound class measure of distance between forms, there is a significant non-linear relationship with the proportion of Norse-derived forms (for statistics, see Table 2 below and section 5.2 in the Data S1). Figure 3 visualises the model's predictions, showing the marginal effects of form distance on the relative proportion of Norse-derived forms. Overall, there is a negative relationship: Norse-derived forms that are very similar to their English counterparts are likely

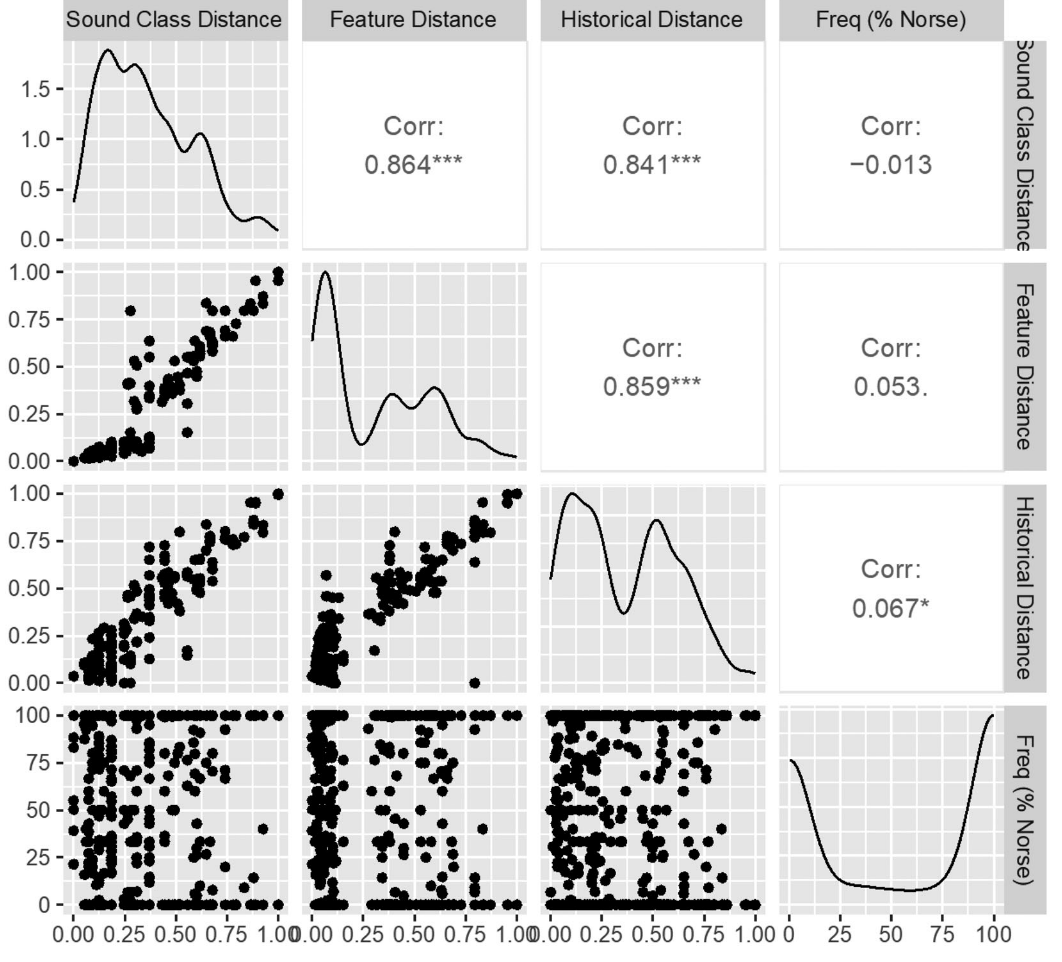

Figure 2. Correlogram statistics for the distance measures and the proportion of attested Norse forms for each pair of comparisons between Norse-derived and English forms. The lower left tringle shows a scatterplot, the top right tringle shows the linear correlation coefficient (with stars indicating significance) and the diagonal shows the univariate density of datapoints for each variable.

Table 2. Statistical results for the total frequency analysis[a]

Total frequency analysis		Model comparison					Model estimate (cubic)		
Measure	Term	BIC	Log likeli-hood difference	χ^2	df	p	Estimate	z	p
Sound Class	Linear	32156	0	0	1	0.885	−10.56	−11.153	<0.001
	Quadratic	32104.9	29.2	58.5	1	<0.001	31.15	13.996	<0.001
	Cubic	31874.5	118.9	237.8	1	<0.001	−24.17	−15.778	<0.001
Feature	Linear	32269.1	8.2	16.5	1	<0.001	−13.12	−13.58	<0.001
	Quadratic	32186	45.2	90.5	1	<0.001	41.4	15.854	<0.001
	Cubic	31888.2	152.6	305.3	1	<0.001	−34.46	−17.797	<0.001
Historical	Linear	32284.2	0.7	1.4	1	0.241	−3.92	−3.982	<0.001
	Quadratic	32172.8	59.4	118.8	1	<0.001	16.25	6.432	<0.001
	Cubic	32112.3	34	67.9	1	<0.001	−15.27	−8.536	<0.001

[a]The major rows show the results for the three different measures, and the minor rows for the linear, quadratic and cubic models. The model comparison columns show the results of model comparison between the model in the given row and the one above (or between the linear model and the null model without a measure of distance). A low p-value indicates that the more complex model has a significantly better fit than the simpler one, taking into account the extra parameters. The model estimate columns show statistics for the estimates within the cubic model. A low p-value indicates that the estimate is significantly different from zero.

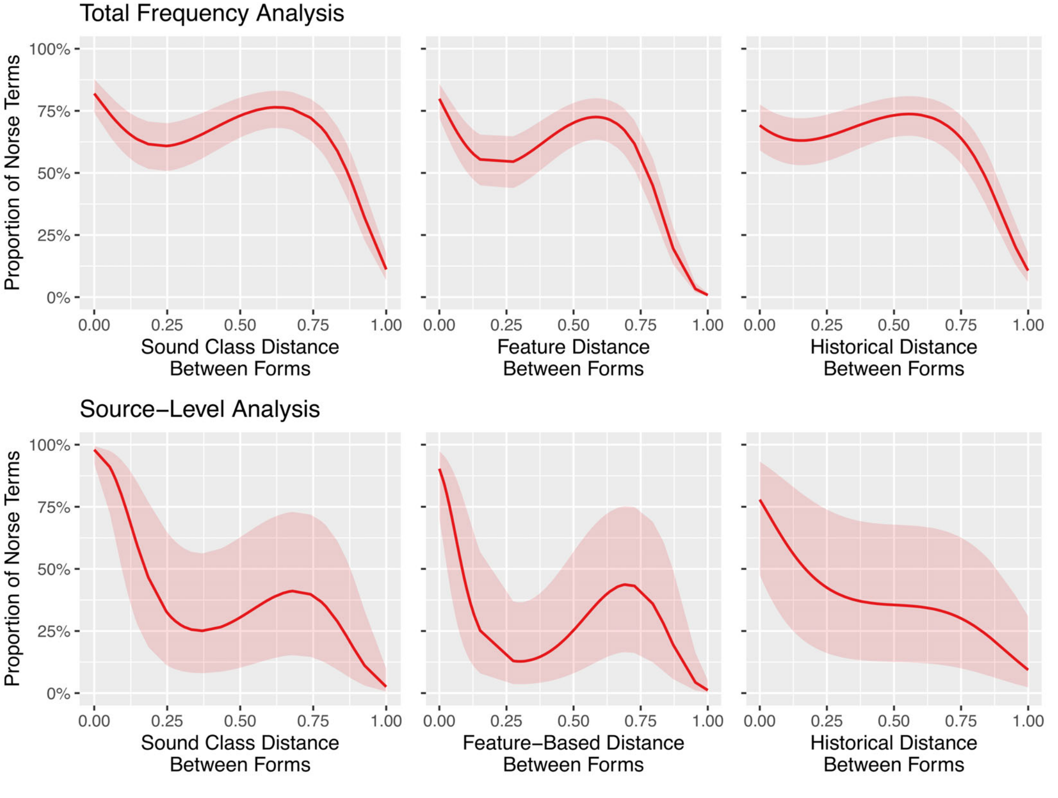

Figure 3. Marginal effects of form distance on the relative proportion of Norse terms. The horizontal axis represents scaled distances between forms from 0.0 (identical, according to the measure) to 1.0 (the greatest distance in the sample). The vertical axis shows the estimated proportion of Norse-derived terms used. Solid lines represent the estimated relationship from the model (some sections are not smooth due to lack of data for those ranges), and the ribbon represents the confidence interval around the estimate.

to have a higher proportion of usage, and Norse-derived forms that are very distant from their English counterparts are likely to have a lower proportion of usage. However, the relationship is not linear because the effect is attenuated for pairs with medium distance (see the curved sections in Figure 2). The same general pattern is found for the feature-based distance measure and the historical distance measure. For the sound class measure and the feature-based measure, the linear effect on its own is not significant, but for every measure the quadratic model is the best fitting model according to the Bayes Information Criteria (or BIC, a number representing the fit of the model, taking into account the number of parameters it uses) and every estimate within each the quadratic model is significant. That is, the non-linear result seems robust across the three measures.

In the source-level analysis, which includes additional controls for source, age and alliteration, the results are broadly similar: a significant negative relationship between form distance and the proportion of Norse-derived forms used. For the sound class measure, there is a significant non-linear relationship that is attenuated for medium ranges (see Figure 3, lower row; for statistics, see Table 3 below and section 6 in the Data S1). The confidence intervals are larger, reflecting the influence of additional controls, but the overall pattern and significance is the same as for the total frequency analysis.[16] The same is true for the feature-based distance and the historical distance. For the historical distance, the quadratic model is only marginally significant, but the cubic model is still the best model and all terms within the cubic model are significant. The Data S1 show the simpler quadratic model, which predicts a very similar relationship to the cubic model. On the relationship between sources and frequency of use, see also below, Section 3.2.

In all the source-level analyses, the effect of alliteration is significant: pairs of forms in alliterative sources that did not alliterate with each other had higher proportions of Norse-derived usage than pairs that did alliterate (e.g. for the sound class analysis, the model estimates that alliterating pairs have a mean Norse proportion of 22% with a very wide range, but pairs that do not alliterate have a mean Norse proportion of 75%, with a more restricted range, $\beta = 2.23$, $z = 31.11$, $p < 0.001$; see section 6 in the Data S1 for more details). That is, Norse-derived forms were more likely to be used if they provided additional flexibility in meeting the requirements of alliterative verse.

Table 3. Statistical results for the source-level analysis[a]

Source-level analysis		Model comparison					Model estimate (cubic)		
Measure	Term	BIC	Log likelihood difference	χ^2	df	p	Estimate	z	p
Sound Class	Linear	20981.2	85.3	170.6	1	<0.001	−32.97	−20.927	<0.001
	Quadratic	20925.7	31.3	62.5	1	<0.001	68.76	19.447	<0.001
	Cubic	20602.4	165.2	330.4	1	<0.001	−43.32	−18.488	<0.001
Feature	Linear	20946.5	102.6	205.3	1	<0.001	−32.43	−18.2	<0.001
	Quadratic	20907.2	23.2	46.3	1	<0.001	77.44	17.015	<0.001
	Cubic	20638.1	138.1	276.2	1	<0.001	−51.69	−16.373	<0.001
Historical	Linear	20949.3	101.3	202.5	1	<0.001	−10.43	−7.388	<0.001
	Quadratic	20951.8	2.3	4.5	1	0.033	19.97	5.476	<0.001
	Cubic	20934.9	12	23.9	1	<0.001	−13.07	−5.176	<0.001

[a]The format in Table 3 is the same as in Table 2 (see Table 2 footnote).

[16] The supporting materials show various other analyses as well, including General Additive Models and a beta-binomial model that can better capture the bimodal distribution of proportions of Norse-derived terms. The results are highly consistent, suggesting a (non-linear) negative relationship between the proportion of Norse-derived terms and form distance.

3.2. *Exploratory analyses*

The results of our exploratory analyses are as follows:

3.2.1. *Word class*

There are no significant differences in the proportion of use of Norse-derived verbs compared to other word classes (model comparison $\chi^2(1) = 0.09$, $p = 0.76$), nor is the model significantly improved by accounting for all word classes (model comparison $\chi^2(6) = 7.06$, $p = 0.31$). Importantly, our results might have been different had we taken into account the specific verbal forms recorded in our corpus (e.g. finite vs. non-finite forms), as this has been shown to make a difference in terms of the integration of Norse-derived verbs in English (see Elter forthcoming, with references).

3.2.2. *Distinctiveness of the diagnostic features*

We have tested whether the model fit is improved when adding a variable that identifies whether the form includes Norse [sk] vs. the native [ʃ] (see above, no. (2) in Section 2.6). Although only twelve pairs have this property, the variable still significantly improves the model (model comparison $\chi^2(1) = 10.24$, $p = 0.001$). The model suggests that, when taking into account cognate set, lexeme and form distance, Norse-derived forms that include this cluster, which is a strong indicator of a Norse origin, have a lower chance of being used than Norse-derived forms that do not (mean of 17% vs. 62%, respectively; $\beta = -2.06$, $z = -3.193$, $p = 0.001$). This fits the general pattern: more distinct Norse words are less likely to be used. The distance measure remains significant in this model. The Data S1 explore other diagnostic velar consonants and find a similar effect. However, the Norse Diagnostic Score does not significantly improve the fit for the source-level analysis, which means that this result is not as robust as the correlation between formal distance and proportion of use.

3.2.3. *Textual differences*

Figure 4 shows the neighbour net comparing the similarity of use of Norse-derived forms between sources. This analysis only considers the proportion of usage, not the distance between forms or any of the other variables in the regression. The distance between two texts is represented by the length of the shortest line that connects them; 'boxes' indicate conflicting signals (e.g. a source being similar for some sets to one group and similar for other sets to another group). Many of the distances reflect geographic distances. For example, the East Midlands sources roughly cluster on the right, the North-West Midlands sources on the left and the Northern sources at the top, closer to documents from Lincolnshire than those from Norfolk. In terms of Northern texts, interestingly, the choices in the *Wars* manuscript are also very close to those in the North-West Midlands texts, which reflects the likely dialectal origin of the text itself (see above, no. (10) in Section 2.1). However, within these groups, distances also reflect diachronic spread. For example, from *FCPC* anti-clockwise to Mannyng's *Chronicle*, the texts are roughly in chronological order,[17] as they are from the *Gawain*-poet's poems clockwise to Rolle's texts. Notably, the local documents from Lincolnshire, Norfolk

[17] As Mannyng was a native of Bourne (see Hanna 2019, with references) and the Petyt manuscript of the *Chronicle* originates from Lincolnshire (see above, no. (5) in Section 2.1), a comparison of the linguistic choices in the *Ormulum*, Mannyng's *Chronicle*, and the Lincolnshire and Nottinghamshire documents might have been expected to offer further insights into the localisation of the *Ormulum* (see above, no. (2) in Section 2.1). However, the results presented in Figure 4 are not conclusive in this respect.

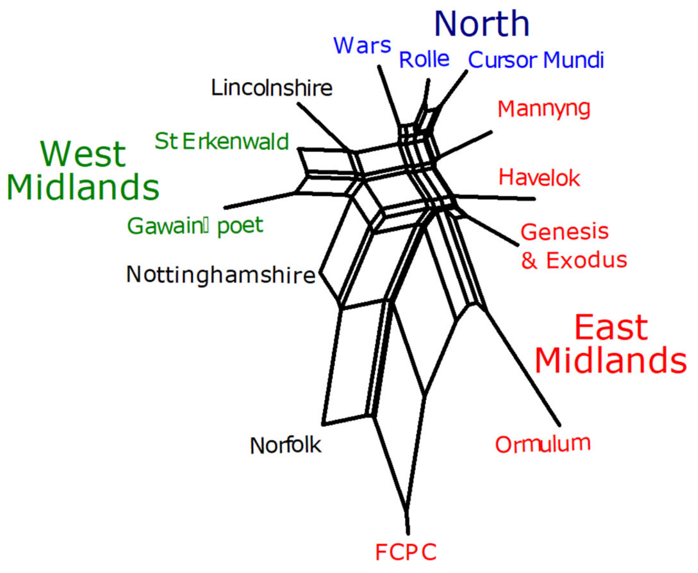

Figure 4. Neighbour net between sources, based on the similarity of the proportion of Norse-derived forms used within each set.

and Nottinghamshire are not placed with their geographic neighbours as, in general, they are closer to texts from the North-West Midlands (although it is interesting to see the proximity between the local documents from Lincolnshire and Mannyng's *Chronicle*). This could be taken as an indication that, in this case, the date of the texts might be more important than their dialectal origin. Regarding the stylistic structure of the texts, there is not an obvious division between prose and poetry, as prose texts (e.g. local documents and Rolle's works) are mixed with the poetic texts (e.g. *Gawain*-poet's compositions, *Erkenwald*, *Wars* and *Cursor*). The situation is not as clear when it comes to different poetic forms, because alliterative and non-alliterative compositions correspond, in the main, to a West–East Midlands divide, but note the closer alignment of *Cursor* with the non-alliterative compositions from the East Midlands (cf. above, Section 3.1).

4. Conclusion

The integration of Norse-derived terms in medieval English is an aspect of early medieval Anglo-Scandinavian linguistic contact that has received much less scholarly attention than the identification of the Norse-derived terms themselves. The latter is a necessary step, but only by studying how and why the Norse-derived terms became integrated into English can we gain a deeper understanding of the make-up of the English lexicon and, more broadly, the way in which language contact can facilitate language change.

We have been able to make a novel contribution in this respect by analysing the relationship between the formal similarity of Norse-derived terms to their native equivalents on the one hand, and their proportion of use in thirteen sources from Scandinavianised areas on the other. Overall, through the implementation of innovative and rigorous approaches to the coding of the data and the measurement of formal distance (see Section 2), we have found strong support for a negative relationship between the similarity of a Norse-derived form to its native equivalent and the frequency of that form's use: the greater the difference, the less frequently Norse-derived forms were used in comparison to the native forms. This finding is, generally speaking, in keeping with the impact that the existence of cognates has on the linguistic output of bilingual speakers and can be used to inform further discussions on the psycholinguistic processes of lexical choice in contact situations involving closely related languages (see Section 1). In terms of the Anglo-Scandinavian contact in particular, it is also fully in line with the fact that Norse-derived verbs which had the same argument structure as their English cognates became more easily integrated into English than those which did not (e.g. Elter 2024 and forthcoming). Moreover, we have explored the interaction between formal similarity and specific stylistic needs, alliteration in particular; we have found that, in poems where alliteration was the main structural device, Norse-derived terms that did not alliterate with their native equivalents had a higher proportion of use than those which alliterated. In this respect, our findings provide much-needed statistical support for commonly rehearsed claims that the similarity between Old English and Old Norse is likely to have facilitated the use of Norse-derived terms in English, and that the requirements of alliteration had a significant impact on (authorial and scribal) lexical choices in texts associated with the Alliterative Revival.

Notably, all statistical analyses indicate a non-linear relationship between form distance and frequency of use because of the results for pairs with medium distance. It is tempting to treat this as evidence of two different psychological effects. First, similarity affects ease of integration: if forms are very similar, then writers hardly notice the choice and forms can become easily integrated into the language. Secondly, forms that have middling differences may actually convey some useful semantic/dialectal/diaphasic difference or nuance, so forms also persist there. However, there are reasons to resist this complex interpretation. Firstly, the results for the historical distance from the source-level analysis had weaker statistical support for a quadratic relationship (suggesting a simpler negative relationship). If the historical distance is the most relevant measure for this historical study, then this is worth paying attention to. Secondly, there are fewer datapoints in the middle range of distances and frequencies, and some of the distance measures have artificial peaks in the distribution due to discrete form lengths. While this is not a violation of the modelling assumptions, it does make the relationship between the variables in the middle of the distribution less certain. Therefore, we argue that the overall picture is that the effect exists for extreme ends of form distance. That is, there may only be one mechanism at work. Future work might collect more data and consider using a Bayesian framework to compare various parameterisations of the model to better understand these issues.

Besides addressing our main hypothesis, we have also conducted a series of exploratory studies around the proportion of Norse-derived terms used in English. We have found some evidence that their rate of use might also be related to how distinctively different the Norse-derived terms were (initially, regarding their Norse origin and later in connection with their association with particular linguistic varieties). However, the effect is not robust for the source-level analysis (perhaps because the latter also controls for other factors such as the age of the text and the stylistic needs of alliterative poetry). Moreover, the direction of the effect requires some explanation because, as far as terms like PDE *again*, *get*, *give* or *scathe* are concerned, these are precisely the forms that have become categoric in Present-Day English. More work needs to be done to investigate this effect, taking into account the interaction between formal similarity, perceived distinctiveness and other factors, such as the existence of homonyms (e.g. *MED* mentions three verbs with the form *yēten* and one under *yĕten*), the semantic nuances that arise, for instance, from differences between cognates in terms of their semantic space, or the stylistic effects that the terms facilitate (consider, for instance, the common alliterative pairing of *scathe* and the French-derived *scorn*; see *Oxford English Dictionary*, s.v. *scathe*, n., sense 2.d).

More broadly, our findings can also inform the study of Middle English dialectology, for the neighbour net in Figure 4 shows clear geographic and temporal relations between our texts in terms of the proportion of use of Norse-derived forms. This suggests that there are important historical signals in the frequency of use of Norse-derived vs. English forms, not just in the choice of the Norse-derived terms themselves. Accordingly, this information can be used alongside other linguistic markers to inform the analysis of texts with uncertain date or place of composition.

All in all, in line with Matthews's thinking, our results attest that, to understand the outcomes of language contact, we must take into account the whole range of factors that are likely to have influenced 'the creative usage of successive groups of [a language's] own speakers' and, in turn, the cultural evolution of specific forms.

Conflicts of interest

The authors do not wish to disclose any conflicts of interest.

Data availability statement

The code and data for this study are available in a GitHub repository (https://github.com/seannyD/IntegrationOfNorseTerms) and archived on OSF (https://doi.org/10.17605/OSF.IO/5F96Z).

Correspondence
Sara M. Pons-Sanz
School of English, Communication and Philosophy
Cardiff University
Cardiff CF10 3EU, UK
Email: pons-sanzs@cardiff.ac.uk

References

[*CMEPV* =] *Corpus of Middle English Prose and Verse* Developed by the University of Michigan Humanities Text Initiative https://quod.lib.umich.edu/c/cme/ (23 December, 2024)

[*MED* =] *Middle English Dictionary*, 2018. Online edition of the *Middle English Dictionary* edited by Robert E. Lewis et al. 1952–2001, in Frances McSparran, et al. (eds.), *Middle English Compendium*. Ann Arbor (MI): University of Michigan Press https://quod.lib.umich.edu/m/middle-english-dictionary/dictionary (10 January, 2025)

A Corpus of Middle English Local Documents, version 2017.1, 2020. Compiled by Merja Stenroos, Kjetil V. Thengs & Geir Bergstrøm University of Stavanger. http://www.uis.no/meld (23 December, 2024)

Ahmed, Tafseer, Muhammad Nizami, Muhammad Yaseem Khan & Alessandro Bogliolo, 2021. 'Discovering lexical similarity using articulatory feature-based phonetic edit distance', *IEEE Access* 10. 1533–1544.

Andrew, Malcolm, 1997. 'Theories of authorship', in Derek Brewer & Jonathan Gibson (eds.), *Companion to the Gawain Poet*. Cambridge: Brewer. 23–33.

Andrew, Malcolm & Ronald Waldron, 2007. *The Poems of the Pearl Manuscript: Pearl, Cleanness, Patience, Sir Gawain and the Green Knight* (5th ed.). Exeter: University of Exeter Press.

Arngart, Olof, 1968. *The Middle English Genesis and Exodus: Re-Edited from MS. C. C. C. C. 44 with Introduction, Notes and Glossary*. Lund: Gleerup.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker, 2015. 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software* 67(1). 1–48.

Benskin, Michael & Margaret Laing, 1981. 'Translations and *Mischsprachen* in Middle English manuscripts', in Michael Benskin & M. L. Samuels (eds.), *So Meny People Longages and Tonges: Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh*. Edinburgh: The Editors. 55–106.

Benson, Larry D., 1965. 'The authorship of *St*. Erkenwald', *Journal of English and Germanic Philology* 64. 393–405.

Borroff, Marie, 2006. 'Narrative artistry in *St. Erkenwald* and the *Gawain*-group: The case for common authorship reconsidered', *Studies in the Age of Chaucer* 28(1). 41–76.

Bosma, Evelyn, Elma Blom, Eric Hoekstra & Arjen Versloot, 2019. 'A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary', *International Journal of Bilingual Education and Bilingualism* 22(4). 371–385.

Burrow, J. A., 1993. '*St Erkenwald* line 1: 'At London in Englond'', *Notes and Queries* 40. 22–25.

Cole, Marcelle, Brian Golding & Eleanor Rye, forthcoming. 'Localizing the Ormulum', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Cooper, Andrew & Johanna Åberg, forthcoming. 'Phonology and orthography in the *Ormulum*', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper, & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Cooper, Andrew & Kees Dekker, forthcoming. 'Manuscripts and editions: A biography of the *Ormulum*', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Croft, William, 2000. *Explaining Language Change: An Evolutionary Approach*. London: Longman.

Dance, Richard, 2003. *Words Derived from Old Norse in Early Middle English: Studies in the Vocabulary of the South-West Midland Texts*. Medieval and Renaissance texts and studies 246. Tempe, AZ: Center for Medieval and Renaissance Studies.

Dance, Richard, 2019. *Words Derived From Old Norse in 'Sir Gawain and the Green Knight': An Etymological Survey*. Transactions of the Philological Society 116, 2 vols. Oxford: Wiley-Blackwell.

Dance, Richard & Sara M. Pons-Sanz, 2025. 'Scandinavian influence', in Laura Wright & Raymond Hickey (eds.), *New Cambridge History of the English Language*, Vol. 1: *Context, Contact and Development*. Cambridge: Cambridge University Press, 219–246.

Dance, Richard, Sara M. Pons-Sanz & Brittany Schorn, 2019. The Gersum Project: The Scandinavian Influence on English Vocabulary, Cambridge, Cardiff and Sheffield https://www.gersum.org (9 January, 2025)

Dance, Richard, Sara M. Pons-Sanz & Brittany Schorn, 2025. 'Introduction: Conversations with Vikings', in Richard Dance, Sara M. Pons-Sanz, & Brittany Schorn (eds.), *The Legacy of Medieval Scandinavian Encounters with England and the Insular World*. Studies in the Early Middle Ages 34. Turnhout: Brepols. 15–43.

Duggan, Hoyt N. & Thorlac Turville-Petre, 1989. *The Wars of Alexander*. Early English Text Society s.s. 10. Oxford: Oxford University Press.

Durkin, Philip, 2014. *Borrowed Words: A History of Loanwords in English*. Oxford: Oxford University Press.

[*eLALME* =], Michael Benskin, Margaret Laing, Vasilis Karaiskos & Keith Williamson, 2024. *An Electronic Version of A Linguistic Atlas of Late Mediaeval English* (Version 2.0). Edinburgh: The Angus McIntosh Centre for Historical Linguistics http://amc.ppls.ed.ac.uk/elalme/ (13 December, 2024)

Elter, W. Juliane, 2024. 'Loan verbs in contact situations between closely related languages strengthening existing argument structural patterns', *Etudes Médiévales Anglaises: A French Journal of English Medieval Studies* 103. 217–253.

Elter, W. Juliane, forthcoming. 'Structural integration of Norse-derived verbs in the *Ormulum*', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper, & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Elter, W. Juliane & Marlieke Shaw, 2025. 'Loan verb accommodation: A comparison of Old Norse and French in Middle English', *English Language and Linguistics* 29(1). 35–58.

Fulk, R. D., 2012. *An Introduction to Middle English: Grammar – Texts*. Peterborough (ON): Broadview Press.

Goering, Nelson, forthcoming. 'Orrm's Prosody', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper, & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Golding, Brian & Eleanor Rye, forthcoming. 'Orrm's world(s)', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper, & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Gooskens, Charlotte, 2024. *Mutual Intelligibility between Closely Related Languages*. Language Contact and Bilingualism 30. Berlin: De Gruyter Mouton.

Gooskens, Charlotte & Femke Swarte, 2017. 'Linguistic and extra-linguistic predictors of mutual intelligibility between Germanic languages', *Nordic Journal of Linguistics* 40(2). 123–147.

Gooskens, Charlotte & Vincent J. Van Heuten, 2020. 'How well can intelligibility of closely related languages in Europe be predicted by linguistic and non-linguistic variables?', *Linguistic Approaches to Bilingualism* 10. 351–379.

Hamming, Richard W., 1950. 'Error detecting and error correcting codes', *The Bell System Technical Journal* 29(2). 147–160.

Hanna, Ralph, 2007. *Richard Rolle: Uncollected Prose and Verse with Related Northern Texts*. Early English Text Society o.s. 329. Oxford: Oxford University Press.

Hanna, Ralph, 2019. 'Robert Manning: Some textual—and biographical—emendations', *Notes and Queries* 66(1). 26–28.

Heeringa, Wilbert, Charlotte Gooskens & Vincent J. Van Heuven, 2023. 'Comparing Germanic, Romance and Slavic: Relationships among linguistic distances', *Lingua* 287. 103512

Herzman, Ronald B., Graham Drake & Eve Salisbury, 1999. *Four Romances of England*. Kalamazoo (MI): Medieval Institute Publications.

Hogg, Richard M., 2011. *A Grammar of Old English, Vol. 1: Phonology*. Oxford: Wiley-Blackwell.

Horrall, Sarah M., 1978. *The Southern Version of Cursor Mundi*. Ottawa: University of Ottawa Press.

Irvine, Susan, 2004. *The Anglo-Saxon Chronicle: A Collaborative Edition*, Vol. 7: *MS E*. Cambridge: Brewer.

Jäger, Gerhard, 2015. 'Support for linguistic macrofamilies from weighted sequence alignment', *Proceedings of the National Academy of Sciences* 112(41). 12752–12757.

Johannesson, Nils-Lennart & Andrew Cooper, 2023. *Ormulum: An Edition from Oxford, Bodleian Library, MS Junius 1, and London, Lambeth Palace Library, MS 731*, Vol. 1: *Text and Glossary*. Early English Text Society o.s. 360. Oxford: Oxford University Press.

Jordan, Richard, 1974, in Eugene Joseph Crook (ed.), *A Handbook of Middle English Grammar: Phonology*, translated and revised by. The Hague: Mouton.

Keller, Jonas, 2020. 'The Leipzig-Jakarta list as a means to test old English/old Norse mutual intelligibility', *NOWELE: North-Western European Language Evolution* 73. 252–275.

Kries, Susanne, 2003. *Skandinavisch-schottische Sprachbeziehungen im Mittelalter: Der altnordische Lehneinfluss*. North-Western European Language Evolution 20. Odense: University Press of Southern Denmark.

[*LAEME* =] Laing, Margaret, 2024. *A Linguistic Atlas of Early Middle English, 1150–1325* (Version 4.0, redesigned by Vasilis Karaiskos). Edinburgh: The Angus McIntosh Centre for Historical Linguistics http://amc.ppls.ed.ac.uk/laeme/ (13 December, 2024)

Laing, Margaret & Roger Lass, 2019. 'Old and Middle English spellings for OE hw-, with special reference to the 'qu-' type: In celebration of LAEME, (e)LALME, LAOS and CoNE: In memoriam Angus McIntosh', in Rhona Alcorn, Joanna Kopaczyk, Bettelou Los, & Benjamin Molineaux (eds.), *Historical Dialectology in the Digital Age*. Edinburgh: Edinburgh University Press. 91–112.

Lamberts, Jacob J., 1953. The dialect of *Cursor Mundi* (Cotton MS Vespasian A III) Doctoral dissertation, University of Michigan. https://deepblue.lib.umich.edu/handle/2027.42/157511 (13 December, 2024)

León Prol, Marta, 2025. Dental instability in northern Middle English: Evidence for lenition and fortition Doctoral dissertation, University of Seville.

List, Johann-Mattis, Cormac Anderson, Tiago Tresoldi, Christoph Rzymski & Robert Forkel, 2024. CLTS: Cross-linguistic transcription systems, version 2.3.0 [Data set], Zenodo. https://doi.org/10.5281/zenodo.10997741

List, Johann-Mattis & Robert Forkel, 2024. LingPy: A Python library for historical linguistics, version 2.6.13 https://lingpy.org, https://doi.org/10.5281/zenodo.10997741

McIntosh, Angus, 1962. 'The textual transmission of the alliterative *Morte Arthure*', in Norman Davis & C. L. Wrenn (eds.), *English and Medieval Studies: Presented to J. R. R. Tolkien on the Occasion of his Seventieth Birthday*. London: George Allen and Unwin Ltd. 231–240.

McMahon, April & Robert McMahon, 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.

Minkova, Donka, 1992. 'Verse structure in the Middle English "Genesis and Exodus"', *Journal of English and Germanic Philology* 91. 157–178.

Monaghan, Padraic & Sean G. Roberts, 2019. 'Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing', *Cognition* 186. 147–158.

Monaghan, Padraic & Sean G. Roberts, 2021. 'Iconicity and diachronic language change', *Cognitive Science* 45(4). e12968

Morris, Richard, 1874–1893. *Cursor Mundi (the Cursur o the World): A Northumbrian Poem of the XIVth Century in Four Versions*. Early English Text Society o.s. 57, 99 and 101. London: K. Paul, Trench, Trübner & co.

Needleman, Saul B. & Christian D. Wunsch, 1970. 'A general method applicable to the search for similarities in the amino acid sequence of two proteins', *Journal of Molecular Biology* 48(3). 443–453.

Nerbonne, John & Wilbert Heeringa, 2010. 'Measuring dialect differences', in Jürgen Erich Schmidt & Peter Auer (eds.), *Language and Space: Theories and Methods*. Berlin: Mouton De Gruyter. 550–566.

Otwinowska, Agnieszka, 2016. *Cognate Vocabulary in Language Acquisition and Use: Attitudes, Awareness, Activation*. Bristol: Multilingual Matters.

Oxford English Dictionary, 2000-. Online edition. Oxford: Oxford University Press. https://www.oed.com (10 January, 2025)

Parsons, David, 2001. 'How long did the Scandinavian language survive in England? Again', in James Graham-Campbell, Richard Hall, Judith Jesch, & David N. Parsons (eds.), *Vikings and the Danelaw: Select Papers from the Proceedings of the Thirteenth Viking Congress, Nottingham and York, 21–30 August 1997*. Oxford: Oxbow Books. 299–312.

Peperkamp, Sharon, 2004. 'A psycholinguistic theory of loanword adaptations', in Marc Ettlinger, Nicholas Fleischer, & Mischa Park-Doob (eds.), *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*. Berkeley (CA): Berkeley Linguistics Society. 341–352.

Perry, George G., 1866. *English Prose Treatises of Richard Rolle de Hampole, Edited from Robert Thornton's Ms. in the Library of Lincoln Cathedral*. Early English Text Society o.s. 20. London: Trübner.

Pons-Sanz, Sara M., 2004. 'A sociolinguistic approach to the Norse-derived words in the glosses to the Lindisfarne and Rushworth gospels', in Christian Kay, Carole Hough, & Irené Wotherspoon (eds.), *New Perspectives on English Historical Linguistics: Selected Papers from 12 ICEHL, Glasgow, 21-26 August 2002. Volume 2: Lexis and Transmission*. Current Issues in Linguistic Theory 252. Amsterdam: John Benjamins. 177–192.

Pons-Sanz, Sara M., 2013. *The Lexical Effects of Anglo-Scandinavian Linguistic Contact on Old English*. Studies in the Early Middle Ages 1. Turnhout: Brepols.

Pons-Sanz, Sara M., 2015. 'Norse-derived terms in Orm's lexico-semantic field of EMOTION', *Journal of English and Germanic Philology* 114(4). 552–586.

Pons-Sanz, Sara M., 2024. *The Norse-Derived Terms in the* Ormulum: *A Reappraisal*. Early Middle English Texts. Leeds: Arc Humanities Press.

Pons-Sanz, Sara M., 2025. 'The lexico-semantic distribution of Norse-derived terms in late Middle English alliterative poems: Analysing the *Gersum* database', in Richard Dance, Sara M. Pons-Sanz, & Brittany Schorn (eds.), *The Legacy of Medieval Scandinavian Encounters with England and the Insular World*. Studies in the Early Middle Ages 34. Turnhout: Brepols. 331–368.

Pons-Sanz, Sara M., forthcoming. 'Orrm's Norse-derived vocabulary', in Sara M. Pons-Sanz, Belén Méndez Naya, Andrew Cooper, & Marcelle Cole (eds.), *The Language of the Ormulum*. Turnhout: Brepols.

Poplack, Shana, Suzanne Robillard, Nathalie Dion & John C. Paolillo, 2020. 'Revisiting phonetic integration in bilingual borrowing', *Language* 96(1). 126–159.

Putter, Ad, 2000. 'Spelling, grammar and metre in the works of the 'Gawain'-poet', *Parergon* 18(1). 77–95.

Putter, Ad, Judith Jefferson & Myra Stokes, 2007. *Studies in the Metre of Alliterative Verse*. Medium Ævum monographs 25. Oxford: Society for the Study of Medieval Languages and Literature.

Putter, Ad & Myra Stokes, 2007. 'The *linguistic atlas* and the dialect of the *Gawain* poems', *Journal of English and Germanic Philology* 106(4). 468–491.

R Core Team, 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing https://www.R-project.org/ (09 January, 2025)

Rogers, James, Stuart Webb & Tatsuya Nakata, 2015. 'Do the cognacy characteristics of loanwords make them more easily learned than noncognates?', *Language Teaching Research* 19(1). 9–27.

Schliep, Klaus Peter, 2011. 'phangorn: phylogenetic analysis in R', *Bioinformatics* 27(4). 592–593.

Skaffari, Janne, 2009. *Studies in Early Middle English Loanwords: Norse and French Influences*. Anglicana Turkuensia 26. Turku: University of Turku.

Skeat, Walter W., 1886. *The Wars of Alexander: An Alliterative Romance Translated Chiefly from the Historia Alexandri Magni de Preliis*. Early English Text Society e.s. 47. London: Trübner.

Smithers, G. V., 1987. *Havelok*. Oxford: Clarendon.

Stenbrenden, Gjertrud F., 2019. 'The development of Old English ǣ: Middle English spelling evidence', in Rhona Alcorn, Joanna Kopaczyk, Bettelou Los, & Benjamin Molineaux (eds.), *Historical Dialectology in the Digital Age*. Edinburgh: Edinburgh University Press. 113–132.

Stenroos, Merja, 2002. 'Words for MAN in the transmission of *Piers Plowman*', in Javier Díaz Vera (ed.), *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics*. Costerus n.s. 141. Leiden: Brill. 375–409.

Sullens, Idelle, 1983. *Robert Mannyng of Brunne: Handlyng Synne*. Binghamton (NY): Center for Medieval and Early Renaissance Studies.

Sullens, Idelle, 1996. *Robert Mannyng of Brunne: The Chronicle*. Binghamton (NY): Center for Medieval and Early Renaissance Studies.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor, 2010. 'Borrowability and the notion of basic vocabulary', *Diachronica* 21. 226–246.

The Middle English Texts Series, 2024. The Rossell Hope Robbins Library at the University of Rochester https://metseditions.org (09 November, 2024)

Thomas, Alexander D. M., 2019. Boundaries, law and governance: The landscape of the Danelaw boundary Doctoral dissertation, University of Bristol. https://research-information.bris.ac.uk/ws/portalfiles/portal/264679537/Final_Copy_21_01_21_Thomas_A_D_M_PhD.pdf (20 January, 2025)

Thompson, John J., 1998. *The Cursor Mundi: Poem, Text and Contexts*. Medium Ævum Monographs n.s. 19. Oxford: Society for the Study of Medieval Languages and Literature.

Townend, Matthew, 2002. *Language and History in Viking Age England: Linguistic Relations between Speakers of Old Norse and Old English*. Studies in the Early Middle Ages 6. Turnhout: Brepols.

Townend, Matthew, 2009. *The Vikings and Victorian Lakeland: The Norse Medievalism of W. G. Collingwood and his Contemporaries*. Kendal: Cumberland and Westmorland Antiquarian and Archaeological Society.

Townend, Matthew, 2024. *The Victorians and English Dialect: Philology, Fiction, and Folklore*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198888123.001.0001

Townend, Matthew, 2025. 'The Vikings and the Victorians and dialect', in Richard Dance, Sara M. Pons-Sanz, & Brittany Schorn (eds.), *The Legacy of Medieval Scandinavian Encounters with England and the Insular World*. Studies in the Early Middle Ages 34. Turnhout: Brepols. 409–439.

Turville-Petre, Thorlac, 2024. *St Erkenwald: A Critical Edition*. Liverpool: Liverpool University Press.

Walkden, George, 2025. 'Scandinavians and verb-second in Northumbrian Old English', in Richard Dance, Sara M. Pons-Sanz, & Brittany Schorn (eds.), *The Legacy of Medieval Scandinavian Encounters with England and the Insular World*. Studies in the Early Middle Ages 34. Turnhout: Brepols. 113–139.

Wełna, Jerzy, 2007. 'The loss of [ei]: [ai] opposition in Middle English', *Studia Anglica Posnaniensia* 43. 3–15.

Wichmann, Søren, Eric W. Holman, Dik Bakker & Cecil H. Brown, 2010. 'Evaluating linguistic distance measures', *Physica A: Statistical Mechanics and its Applications* 389(17). 3632–3639.

Wichmann, Søren, Eric W. Holman & Cecil H. Brown, 2022. *The ASJP Database*, version 20 https://doi.org/10.5281/zenodo.7079637

Winford, Donald, 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.

Wood, Simon N., 2011. 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models', *Journal of the Royal Statistical Society*. *Series B, Statistical Methodology* 73(1). 3–36.

Appendix A
Pairs of Norse-derived and native terms considered in this study

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
1. 'against; again, in return'	Forms with a velar consonant (cf. OIc *(í) gegn*; cf. ME *ayēn*) – A1*c	ME *ayēn* (< OE *ongegn*, *ongēan*)	ME *ayēnes* (adv.) ME *ayēn(e)s* (prep.) ME *ayēnstō̆nden* ME *ayēnward* ME *onyē̆n* ME *thē̆ryēnes* ME *yēncome* ME *yēnsaue*
2. 'always, ever'	ME *ai* (cf. OIc *ei*) – A1*c	ME *ō* (< OE *ā*)	
3. 'cattle; inheritance'	ME <arrf> (cf. OIc *arfr*; cf. ME *erve*) – B2ac[a]	ME *erve* (< OE *ærfe*, *yrfe*)	ME *erveblōd* ME *ervenāme* ME *ervward* ME *ervwardriche*
4. 'fear, dread, awe'	ME *aue* (cf. OIc *agi*) – C1c	ME *eie* (< OE *ege*)	ME *aueful* ME *au(e)lēs* ME *auelī* ME *eielēs*
5. 'blast, gust of wind'	ME *blast* (cf. OIc *blástr*) – C2	<blest> (< OE *blǣst*)	
6. 'cord, string, bond'	ME *bō̆nd* (cf. OIc *band*) – C1ac	ME *bē̆nd(e* (< OE *bend*)	
7. 'both'	ME *bōth(e* (cf. OIc *báðir*) – C4a	ME *bō* (<ME *bēgen*, *bā*, *bū*)	
8. 'pale, pallid'	ME *bleik* (cf. OIc *bleikr*) – A1*	ME *blōk* (< OE *blāc*)	
9. 'flower, bloom'	ME *blōm* (cf. OIc *blómi*, *blóm*) – C3ac	ME *blosme* (< OE *blōstm(a)*)	
10. 'boon, request'	ME *bōn* (cf. OIc *bón*) – C1	ME *bēne* (< OE *bēn*)	

Appendix A: (continued)

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
11. 'to rush, move quickly'	ME *breiþen* (cf. OIc *bregða*) – C2c	ME *breiden* (< OE *bregdan*)[b]	ME *abreiden* ME *outbreiden* ME *upbreiden*
12. 'to burn'	ME *brennen* (cf. OIc *brenna*) – C2c	Forms with metathesis (< OE *beornan*, *biernan*, *byrnan*; *bærnan*; cf. ME *brennen*)	ME *brennere* ME *brenninge* ME *forbrennen*
13. 'to burst, break, shatter'	ME *bresten* (cf. OIc *bresta*) – C2	Forms with metathesis (< OE *berstan*; cf. ME *bresten*)	ME *atbresten* ME *outbresten* ME *tō̆bresten*
14. 'emperor'	ME *caiser* (cf. OIc *keisari*) – C2	ME *cāsere* (< OE *cāsere*)	ME *cāserking*
15. '(lower class, common) man'	ME *carl* (cf. OIc *karl*) – B2a	ME *cherl* (< OE *ceorl*)	ME *carlman*
16. 'to drown'	ME *drounen* (cf. OIc *drukna*; Da. *drunkne*, *droune*) – C2	ME *dronknen* (< OE *druncnian*)	ME *ofdruncnen*
17. 'eighth'	ME *eightend* (cf. OIc *áttandi*) – C2ac	ME *eightethe* (< OE *eahteoða*)	
18. 'few'	ME *fō̆* (cf. OIc *fár*) – A1*c	ME *feue* (< OE *fēawe*)	ME *unfeue*
19. 'to get, acquire'	ME *gĕ̄ten* (cf. OIc *geta*) – A1*c	Forms with initial palatal consonant (< OE WS -*gietan*, -*gitan*, -*gytan*; A -*getan*; cf. ME *gĕ̄ten*)	ME *biyĕ̄ten* ME *foryĕ̄ten* ME *foryĕ̄ting* ME *gĕ̄ting(e*
20. 'to give'	Forms with initial velar consonant (cf. OIc *gefa*; cf. ME *yēven*) – A1*c	ME *yēven* (< OE WS *giefan*, *gyfan*; A *gefan*, *gifan*)	ME *ayēven* ME *foryēven* ME *foryēvenes(se* ME *yēving(e*

Appendix A: (continued)

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
21. ‘gift’	Forms with initial velar consonant (cf. OIc *gefa*; cf. ME *yēve*) – A1*c	ME *yēve* (< OE *gifu*, *giefu*, *geofu*, *geafu*)	
22. ‘gift’	Forms with initial velar consonant (cf. OIc *gift*; cf. ME *yift(e*) A1*c	ME *yift(e* (< OE *gift*, *gyft*)	
23. ‘attention, care’	ME *gōme* (cf. OIc *gaum(r)*) – A1*ac	ME *yēme* (< OE *gīme*, *gȳme*, *gīeme*)	ME *yēmelēst*
24. ‘ready’	ME *greith* (cf. OIc *greiðr*) – A1*c	ME *rēdī* (< OE *gerǣde* + ME -*ī̆*)	ME *greithlī* ME *rēdīlī*
25. ‘steward, reeve’	ME *greive* (cf. OIc *greifi*) – C2ac	ME *rēve* (< OE *gerēfa*)	ME *shī̆rrēve*
26. ‘hail; healthy’	ME *heil* (cf. OIc *heill*) – A1*	ME *hōl(e* (< OE *hāl*)	ME *drink heil* ME *heilnesse* ME *hōllī* ME *hōlsom*[c] ME *hōlsomlī* ME *unhōl(e* ME *unhōlsom* ME *wassail*
27. ‘health; success’	ME *heil* (cf. OIc *heil*) – A1*	ME *hēle* (< OE *hǣl*, *hǣlu*)	ME *hēleful* ME *ille hail* ME *ille hēle* ME *unhēle* ME *wrōtherhēle*
28. ‘to greet’	ME *heilsen* (cf. OIc *heilsa*) – A1*c	ME *halsen* (< OE *hālsian*)	ME *halsing(e* ME *heilsing(e*
	ME *hēthen* (cf. OIc *heðan*) – B1c	ME *henne* (< OE *heonan*)	

Appendix A: (continued)

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
29. ‘hence, from this place’			ME *hennes* ME *hennesfŏ̄rth* ME *hennesfŏ̄rthwărd* ME *hēthenwărd*
30. ‘to cover, conceal, hide’	ME *hilen* (cf. OIc *hylja*) – C1a	ME *hēlen* (< OE *helan*)	ME *forhēlen* ME *unhēlen* ME *unhilen*
31. ‘hair’	Forms with <a>/<o> (cf. OIc *hár*; cf. ME *hēr*) – A1*c	ME *hēr* (< OE WS *hǣr*; A *hēr*)	
32. ‘hundred’	Forms with a final fricative consonant (cf. OIc *hundrað*; cf. ME *hundred*) – A1*c	ME *hundred* (< OE *hundred*)	ME *hundredfōld*
33. ‘church’	Forms with velars (cf. OIc *kirkja*; cf. ME *chirche*) – A1*bc	ME *chirche* (< OE *cyrice*)	ME *chirchedōre* ME *chircheflōr* ME *chirchegăng* ME *chirchewardein* ME *chirchewīk(e* ME *chircheyĕrd*
34. ‘chest, casket’	Forms with velar (cf. OIc *kista*; cf. ME *chest(e)*) – A1*c	ME *chest(e* (< OE *cist*, *cyst*)	
35. ‘loathsome; hateful’	Forms with diphthong (cf. OIc *leiðr*; cf. ME *lōth*) – A1*c	ME *lōth* (< OE *lāth*)	ME *lōthlī* (adj. and adv.) ME *lōthsŏmnes(se*
36. ‘air, sky’	ME *loft* (cf. OIc *loft*, *lopt*) – C1b	ME *lift* (< OE *lyft*)	
37. ‘to loosen’	ME *lōsen* (cf. OIc *lauss*) – A1*	ME *lēsen* (< OE *lēasan*)	ME *forlēsen* ME *lēsing(e* ME *lōsing(e* ME *outlēsen*

Appendix A: (continued)

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
38. 'woman, maiden'	ME *mai* (cf. OIc *mær*, *meyja*) – A1*c[d]	ME *mei* (< OE *mǣg*)	
39. 'in the middle'	ME *emel(le* (cf. ODan. *(i) mellæ*, OSwed. *(i) mälle*) – A1*c	ME *middel* (cf. OE *middel*)	ME *amidward*
40. 'human, manly'	Forms with *-sk* cluster (cf. OIc *mennskr*; cf. ME *mannish*) – C2c	ME *mannish* (< OE *mennisc*)	ME *mannishnes(se* ME *menniscleȝc*
41. 'mouth'	ME *mum* (cf. OIc *munnr*)[e] – A1*bc	ME *mouth* (< OE *mūþ*)	
42. 'no'	ME *nai* (cf. OIc *nei*) – A1*	ME *nō* (< OE *nā*)	
43. 'to use, enjoy'	ME *naiten* (cf. OIc *neyta*) – A1ac	ME *nōten* (< OE *nēotan*)	
44. 'to name, call, mention'	ME *nevenen* (cf. OIc *nefna*) – A1*c	ME *nemnen* (< OE *nemnan*)	ME *unnevened*
45. 'cattle, bull, ox'	ME *nout(e* (cf. OIc *naut*) – A1*c	ME *nēt* (< OE *nēat*)	
46. 'forward movement, course; rush'	ME *rās(se* (cf. OIc *rás*, *ras*) – C2	ME *rēs* (< OE *rǣs*)	
47. 'advice, counsel'	ME *rāth* (cf. OIc *ráð*) – A1*	ME *rēd* (< OE *rǣd*)	ME *orrāth* ME *orrāthnesse* ME *rēdlē̆s* ME *unrēd*
48. 'to advise'	ME *rōthen* (cf. OIc *ráða*) – A1*c	ME *rēden* (< OE WS *rǣdan*, A *rēdan*)	
49. 'to raise, lift up, elevate'	ME *raisen* (cf. OIc *reisa*) – A1*c	ME *rēren* (< OE *rǣran*)	ME *arēren* ME *rēring(e* ME *overraisen* ME *upreisen*

Appendix A: (continued)

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
50. 'harm, injury; disaster'	ME *scā̆th(e* (cf. OIc *skaði*) – A1*C	ME *shāthe* (< OE *sceaþa*)	ME *scāthlḗ̆s* ME *scā̆thlī* ME *unskā̆theful* ME *unskāthelī*
51. 'sister'	<i> forms (cf. OIc *systir*; cf. ME *suster*) – A1*c	ME *suster* (< OE *swustor*, *sweostor*)	ME *halfsuster*
52. 'bright, pure'	ME *skīr(e* (cf. OIc *skírr*) – A1*c	ME *shīr(e* (< OE *scīr*)	
53. 'star'	ME *sterne* (cf. OIc *stjarna*) – B2c	ME *sterre* (< OE *steorra*)	ME *sēsterre* ME *sterneleme*
54. 'from there, thence'	ME *thē̆then* (cf. OIc *þaðan*, *þeðan*) – B1c	ME *thenne* (< OE *þanon*)	ME *thennes* ME *thē̆thenforth* ME *thē̆thenward*
55. 'though, although; however'	ME *though* (cf. OIc *þó*) – A1*c	Forms with <e, a> (< OE *þēah*; cf. ME *though*)	ME *although* ME *thoughwhether*
56. 'news, message'	Forms with a fricative (cf. OIc *tíðindi*; cf. ME *tīding(e)*) – A1*	ME *tīding(e* (< OE *tīdung*)	
57. 'trusty; trustworthy; certain, sure'	ME *traist(e* (cf. OIc *treystr*, *traustr*) – A1c	ME *trust* (<ME *trust*)[f]	
58. 'trust, confidence'	ME *traist(e* (cf. OIc *treysta*) – A1c	ME *trust* (< OE **trust*)	ME *trustī*
59. 'to trust, have confidence in; believe'	ME *traisten* (cf. OIc *treysta*) – A1c	ME *trusten* (< OE **trystan*)[g]	

Appendix A: (continued)

Meaning	Norse term and *Gersum* classification	English term	Other members of their word-fields taken into consideration
60. 'trustworthy, trusty'	ME *trig* (cf. OIc *tryggr*) – A1*c	ME *treu(e* (< OE *trēowe*)	ME *ortroue* ME *treulī* ME *untrue(e*
61. 'from wherever, whence'	ME *whĕthen* (cf. OIc *hvaðan*) – B1c	ME *whenne* (< OE *hwanon*)	ME *whennnes* M *whĕthenward*
62. 'wholly, utterly'	ME *thwertout* (cf. OIc *þvert*) – A3*c	ME *thurghout* (< OE *þurhūt*)	ME *thurghoutlī*
63. 'misery, trouble, woe'	ME *wei* (cf. OIc *vei*) – A1*c	ME *wō* (< OE *wā*)	ME *wēmōd*
64. 'weak'	ME *weik* (cf. OIc *veikr*) – A1*	ME *wōk* (< OE *wāc*)	
65. 'care, keeping, protection'	ME *yēmsle* (cf. OIc *geymsla*) – C1c	ME *yēming* (cf. ME *yēmen*)	

[a] The compound <arrfname> in the *Ormulum* is classified as A1c in Pons-Sanz (2024: 13) on the basis of <name>, rather than <arrf>.

[b] On the possibility that the relevant meaning for the native word-field might be Norse-derived, see Dance et al. (2019: s.v. *brayd*); the noun (ME *breid*) is classified as CC3. Even if the terms represented semantic loans, ME *breiden* has been considered a suitable alternative for ME *breiþen* because its form is native.

[c] On the possibility that this adjective and ME *hōlsomlī* might be Norse-derived, see Dance et al. (2019: s.v. *holsumly*), where the adverb is classified as CCC1a.

[d] This classification applies to the form <maȝȝ> in the *Ormulum*; as noted by Dance et al. (2019: s.v. *may*), in other contexts it is not possible to differentiate the Norse-derived and native forms, hence their classification of the term as CC5b.

[e] <mum> is the spelling provided by *MED* for this headword although, given its suggested etymon and its attested spellings, <mun> would have been a better option (cf. *Oxford English Dictionary* 2000–: s.v. *mun*, n.1).

[f] For a discussion of the possibility that this adjective and ME *trust* (n.) might also be loans based on the same Norse root as ME *traist(e*, see Dance et al. (2019: s.v. *trysty*); the term is classified as BBB2.

[g] For a discussion of the possibility that this verb might also be a loan based on the same Norse verb as ME *traisten*, see Dance et al. (2019: s.v. *tryst*); as is the case for nominal/adjectival forms with *trist-*/*trust-*, the verb is classified as BBB2.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Supporting materials (full statistical report) for the Integration of Norse-Derived terms in English.