Edinburgh Research Explorer

A grapho-phonologically parsed corpus of medieval Scots

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Corpora

A grapho-phonologically parsed corpus of medieval Scots:
Variation across time

ABSTRACT
This paper presents key aspects of the data, methods and uses of the *From Inglis to Scots (FITS) Corpus* (Alcorn *et al.,* 2021–), complementing Kopaczyk (2018) and focusing on the diachronic dimension of the resource. The corpus reconstructs sound values for individual Older Scots morphological root elements of Germanic origin as attested in the documentary record in *Linguistic Atlas of Older Scots* (LAOS). This is done by triangulating between the attested spelling, the sound values proposed in the literature for their etymological sources (dialects of Old English, Old Norse and Middle Dutch), and a series of plausible sound changes leading from the latter to the former (the Corpus of Changes). The challenges and possibilities of this approach are highlighted throughout, focusing on the diachronic mapping of Older Scots sounds to their origins and the intervening changes. An overview of the corpus' capabilities is provided in tandem with its limitations.

1. Introduction

The past twenty years have seen a sharp rise in the use of corpus methods in the study of historical language forms and functions. However, much of this focus has been on generating specific lexical, morphological, semantic or syntactic parsings of historical texts, whilst relatively little attention has been paid to the parsing of sound structure.[1] This said, the close analysis of variation in non-standard spelling systems has always been a key element in the traditional historical-phonological toolkit. The resource we present here aims to bring together these two traditions: the digital, quantitative analysis of historical texts and the meticulous uncovering of sound systems through interpretation of spelling variation. The result is the first corpus of non-standard spellings mapped on to reconstructed sound values: *From Inglis to Scots: A Corpus of Grapho-phonological Correspondences (1380-1500) with Associated Corpus of Changes* (*FITS* – Alcorn *et al.,* 2021–).

Our focus is on Older Scots (OSc), the insular West-Germanic language which emerged as the *lingua franca* of the Scottish burghs in the 12th century, and was originally referred to as *Inglis* and later as *Scots* (see McClure, 1981: 52). As is apparent from the early designation, Scots is a close relative of English and a descendant of northern Old English (OE) dialects, influenced particularly by Old Norse (ON), French and Latin. The FITS corpus focuses on some of the earliest attested writing in the language and tries to explicitly reconstruct its phonic structure and map this — via sound changes — to its etymological sources.

Text production for pre-standard vernacular writing tends to be highly localised, often leading to the coexistence of multiple spelling conventions which draw from diverse orthographic traditions. The FITS corpus analyses the graphemes recorded in Osc documents and offers our best reconstructions for their corresponding sound values. This makes it possible to study both sounds and spellings with the quantitative precision of corpus linguistics, taking into account variables such as linguistic context, time and space (see Alcorn *et al*., 2017; Maguire *et al*., 2019, Molineaux *et al.,* 2019, 2020, 2021).

---

[1] Notable exceptions are the *Eighteenth-Century English Phonology Database* (ECEP) and the *Corpus of Narrative Etymologies* (CoNE).

The approach to explicitly tagging historical phonological data in a corpus, which is laid out here, builds on the notion of *grapho-phonological parsing*, described in Kopaczyk *et al.* (2018). While that paper details the reconstruction of the synchronic sound-spelling relationships for OSc underlying FITS, this article focuses on the historical sources of those reconstructed sounds and how they came by their attested spellings. This entails finding a likely etymological source for each item in the corpus and reconstructing its changes through time, thus proposing a corpus-based approach to phonological change (see Figure 1).
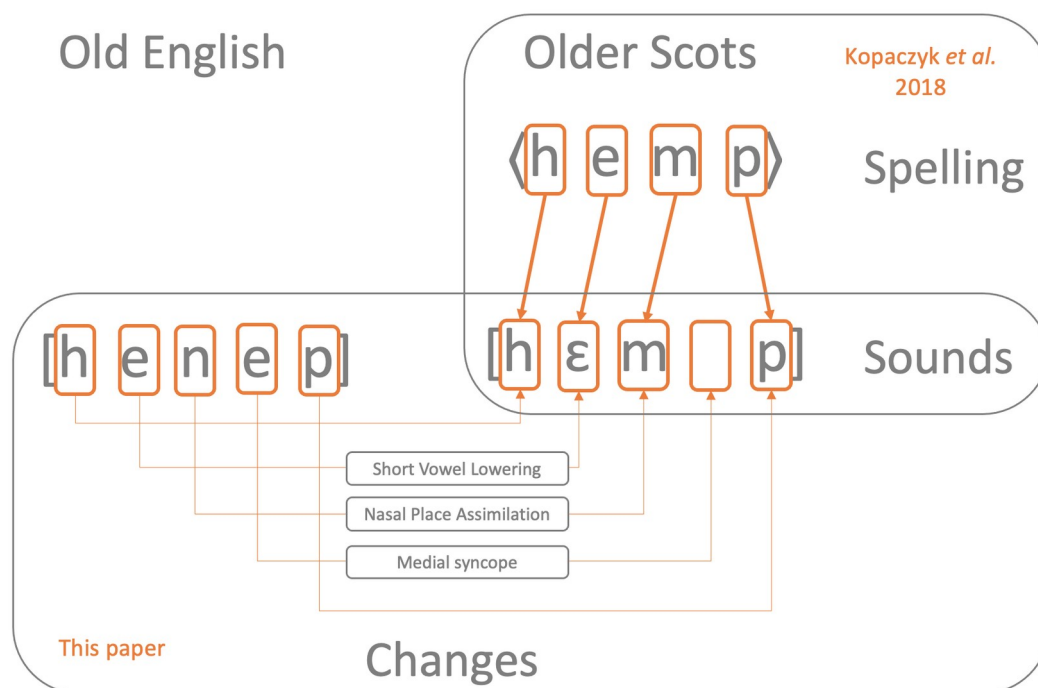

Figure 1: A schematic representation of the entry for 'hemp' in FITS, displaying the Older Scots sound-spelling mappings and the etymological (Old English) sounds which map onto them via sound changes

In what follows, we give an overview of the FITS data, with an emphasis on the proposed etymological sources for the attested OSc forms (§2). We go on to describe the sets of relations between sources, attested OSc spellings and reconstructed sounds, and how we link these together via a Corpus of Changes (§3). §4 lays out the structure of the database itself and presents some key aspects of the front end's search capabilities. We conclude with some prospects for the corpus, alongside a discussion of the methodology's replicability and limitations (§5).

2. The data
The FITS corpus focuses exclusively on the Germanic root morphemes attested in content words from the *Linguistic Atlas of Older Scots* (LAOS – Williamson, 2008), which includes c.1,250 Scottish local documents (c.0.4 million words) spanning the period from 1380 to 1500. The LAOS texts are diplomatically transcribed from manuscripts and semantico-grammatically tagged, allowing us to identify individual roots and their myriad spellings. The texts also contain detailed metadata which usually allow us to pinpoint the texts' geographical and temporal origins. The total number of target morpheme types (based on the LAOS lemmas) is just under 1,100,

spread across c.16,500 word types. Finally, the total token count for all the target morphemes is nearly 110,000.

Since this paper focuses on describing the diachrony of the reconstructed sound values of mediaeval Scots, we must first identify the specific ancestral language from which each item descends. Interestingly, in a survey of entries in the *Dictionary of the Older Scots Tongue* (DOST) Macafee and Anderson (1997) report that only 45.5% of the surveyed entries are of Germanic origin. Indeed, the OSc lexicon is shown to be very rich in words of Latin and Old French origin (46.7%), which is particularly true for legal and administrative texts such as those found in LAOS. In contrast, etymologically Celtic items are quite rare (0.8%).

As a rule, we include in FITS only items which were uncontroversially of Germanic origin, following the etymologies in CoNE, DOST and the OED. The following gives a brief description of the Germanic source languages used to tag FITS root morphemes, detailing the practical decisions involved.

Old Northumbrian: The emergence of OSc is considered to go hand-in-hand with the twelfth-century establishment of Scottish royal burghs, which brought together speakers of several Germanic languages and dialects, alongside Gaelic and Old French speakers (cf. Macafee, 2002; Alcorn *et al*., 2017). The *lingua franca* of this melting pot would have taken on features from these varieties in a process that has been described as *koineisation* (Johnston, 1997a: 56, Millar, 2020). Amongst these sources, the ones with the best claim to represent a *founder* language are the descendants of the local dialect of Old English (OE – Millar, 2020:65). Indeed, OE varieties were spoken in the Scottish Lowlands as far back as the fifth and sixth centuries (Proudfoot and Aliaga-Kelly, 1996), which corresponds to the northernmost kingdom of the Angles, Bernicia, between the rivers Tees and Forth. No written Bernician record survives, but we do have a body of texts in a variety known as Old Northumbrian (ONhb). This represents the language of the early Kingdom of Northumbria, which emerged in the seventh century as the outcome of a dynastic merger between Bernicia and its southern neighbour, Deira. The, admittedly sparse, record for ONhb represents some of the earliest OE writing anywhere, containing *Caedmon's Hymn, Bede's Death Song*, the *Leiden Riddle*, and the *Lindisfarne Gospels* and *Durham Ritual* glosses. While ONhb was probably closely related to the Anglian variety spoken in Lowland Scotland, practically all of these preserved materials originate south of the traditional Scottish border.[2] The ONhb record, furthermore, does not reach beyond the mid tenth century, which means there is also a temporal gap of a full four centuries between the final ONhb materials and the earliest OSc ones (1375).

Given the key role of this variety in the development of OSc, we made a decision to use the 'ONhb' tag in FITS whenever a likely source item (i.e. a source which could give rise to the OSc form via otherwise attested or typologically plausible sound changes) was found in the available materials for the variety. Searches of said materials were conducted on a simple, bespoke TXT corpus, compiled from the *Dictionary of Old English Web Corpus* (DOEWC).

---

[2] The Ruthwell Cross inscriptions (early 8th c. Dumfriesshire) are the main exception.

Anglian: A second DOEWC-based corpus was also compiled for OE texts preserved in the Mercian dialect. Since both ONhb and Mercian are considered Anglian varieties, which differ somewhat from better-documented late West Saxon (cf. Hogg, 2006), we use the 'Anglian' tag to represent this broader grouping and use it to mark words found in the relevant corpus, but missing from the ONhb materials. The Mercian corpus includes texts like the *Corpus* glossary or the *Epinal Glossary*, which also contain more everyday vocabulary, as compared to the predominantly biblical ONhb materials. Even though these Mercian sources suffer from the same time-gap to OSc, they broaden the lexical range of sources in Anglian dialects, functioning as a proxy for ONhb/Bernician. What is more, we also assume that many speakers of the reflexes of these varieties would have been among the migrants to the newly-established Scottish burghs: peasants, traders and craftsfolk who followed the new Norman nobility to Scotland (Barrow, 2003).

Old English: Where a clearly West Germanic morpheme is found which is either impossible to trace to the previous two categories, or fits a late West Saxon phonological shape and changes best, we tag it as 'OE' (Old English). In the very rare cases where ONhb and/or Anglian forms are attested, but a West Saxon form was found to be a better fit, this was noted in the comments to the source (e.g. OE *geornan* [jeorn-ɑn] instead of ONhb *giornan* [jiorn-ɑn] as a source for OSc *yharnand* [jarn-and] 'yearn/PRES.PART'). Such cases are, presumably, the result either of early anglicisation, or of the peculiarities of the scribe's dialect and training.

Old Norse: As it was located above the Scandinavian Belt (Samuels, 1985: 269), the ONhb/Bernician substrate of Scots is not traditionally considered to have as much ON influenced as other dialects of Middle English. Nevertheless, Norse influence on Osc is considerable, being alternatively attributed to Scandinavianised-Anglian immigrants who followed the Anglo-Norman nobility (cf. Johnston, 1997a; Macafee and Aitken, 2002) or from direct influence from Viking settlements in Scotland (Kries, 2007). 'ON' was used as a tag when a morpheme was more clearly derivable from a Norse source than one of the dialects of OE. Considering the close genetic relationship of OE and ON, there are instances were both could plausibly have been the source (cf. Durkin, 2014: 190). In these cases, the one that was phonologically more plausible was used as the source (cf. ON *treist* [treist] vs OE **tryst* [tryst] for OSc *traist* [traist] 'trust') or, where this was impossible, we provide two parallel etymologies, such as in the case of OSc *wecht* [wɛxt] 'weight', plausibly be derived from either ON [wext] or OE [wiçt].

Middle Dutch: There are also a handful of words borrowed through North Sea trade networks and small Flemish settlements in Scotland. Typically, these include shipping, administrative and (cloth) trade vocabulary (Murison, 1971; Macafee and Anderson, 1997) and are tagged as 'MDu'.

The two most common sources across the corpus are ONhb and OE (see Table 1). The type count depends on whether attestations are counted as a single morpheme in the context of a particular grammatical category (GC) or as a unique form. For instance, the root *help* is found under five different GCs: singular and a plural noun (e.g. <helpar; helper(is)>) and simple present, past participle and present participle verbs (e.g. <helpe; helpit; helpand>). When counting by the number of distinct GCs, ONhb has the highest count; when counting by the number of distinct root morphemes, OE

is most common. This represents the fact that low frequency words are less likely to be attested in the limited ONhb corpus.

Source language tag	Morphemes: by GC	Morphemes: unique
ONhb	1362	422
OE	894	432
ON	376	134
Anglian	146	65
MD	32	22

Table 1: Number of morphemes by source language

3. Grapho-phonological parsing and variation across time

Reconstructing the sound values behind spelling forms implies triangulating on a number of factors. Spelling is one of these factors, but the etymological source of a given word and its present-day pronunciations must also be considered, as well as the sound and spelling changes that are likely — or unlikely — to occur in the history of the language. For every instance of a grapheme, therefore, FITS provides a triad made up of (i) a Pre-Scots, etymological phone, (ii) an Older Scots phone and (iii) an Older Scots grapheme. Where there is a discrepancy between the Pre-Scots and the OSc phones, the relevant linguistic development is reconstructed. In the example in Figure 2, the proposed change from [oː] to [øː] is characterised as an instance of NORTHERN FRONTING (NF).[3] Ultimately, FITS can be seen as a corpus of such triads alongside a corpus of changes.
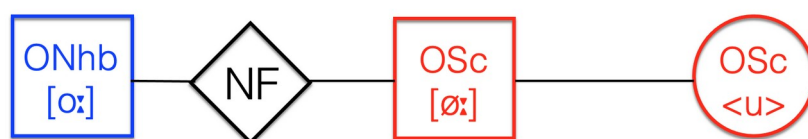

Figure 2: A diachronic grapho-phonological triad

The actual spellings used to represent the outcome of NF can vary considerably. For example Aitken and Macafee (2002: 160) give <o, oi, oy, u, w, oo, ou, ow> all as spellings for the fronted vowel. In the case of OSc <u>, the pattern is attested (among others) in the OSc item <mudir> 'mother', traced back to the ONhb form <modir>. The relevant FITS database entry, linking these three elements via changes can be viewed in Figure 3.

---

[3] Throughout, small capitals represent individual changes. Graphemic representations are given in angled brackets <...>. Scribal abbreviations are expanded in parentheses (...). Sound values are given in IPA using square brackets [...] (cf. Alcorn *et al.*, 2017).

Figure 3: FITS entry for the OSc spelling <mudir> 'mother'.[4]

Each of the ~100 individual proposed processes that link source and OSc sounds is included as a brief narrative in a *Corpus of Changes*. These descriptions (see Figure 4) survey the existing literature on Scots and English, as well as the general typological likelihood of the change. Present-day Scots pronunciations are also considered in shaping our reconstructions,[5] however, we do not attempt to provide a present-day reference form, since there tends to be substantial variation across dialects today.

Figure 4. *Corpus of Changes* entry for NORTHERN FRONTING (NF), with a subset of affected forms[6]

4. Database structure, searches and outputs

The FITS database was created on a MySql server and can be summarised schematically as in Figure 5. The tables under *Synchronic* are described in more detail in Kopaczyk et al. (2018), but represent the LAOS-copied OSc forms (*forms*), each

---

one of which is subsumed under a morpheme (*morphemes*);[7] and split up into spelling units (*graphemes*), and their respective sound values (*sound values*).

The linking of the LAOS forms with their sources is stored in the database tables under *Diachronic*. These source elements are analysed in the tables *sources* and *sourceParts*; which are the pre-Scots equivalents of *forms* and *sound values,* respectively: *sources* stores information on the proposed etymological form and the ID of the source language, *and sourceParts* splits the etymological form into its constituent reconstructed sound values. The source languages are listed in the table *languages*. *Etymologies* connects LAOS forms with their historic forms.

All the changes are described in *changes* and *changeTypes*. In this table we link the sound values in a LAOS form (*sound values*) with their corresponding sound values in historic forms (*sourceParts*), and the sound changes that occur in the transition from the latter to the former.
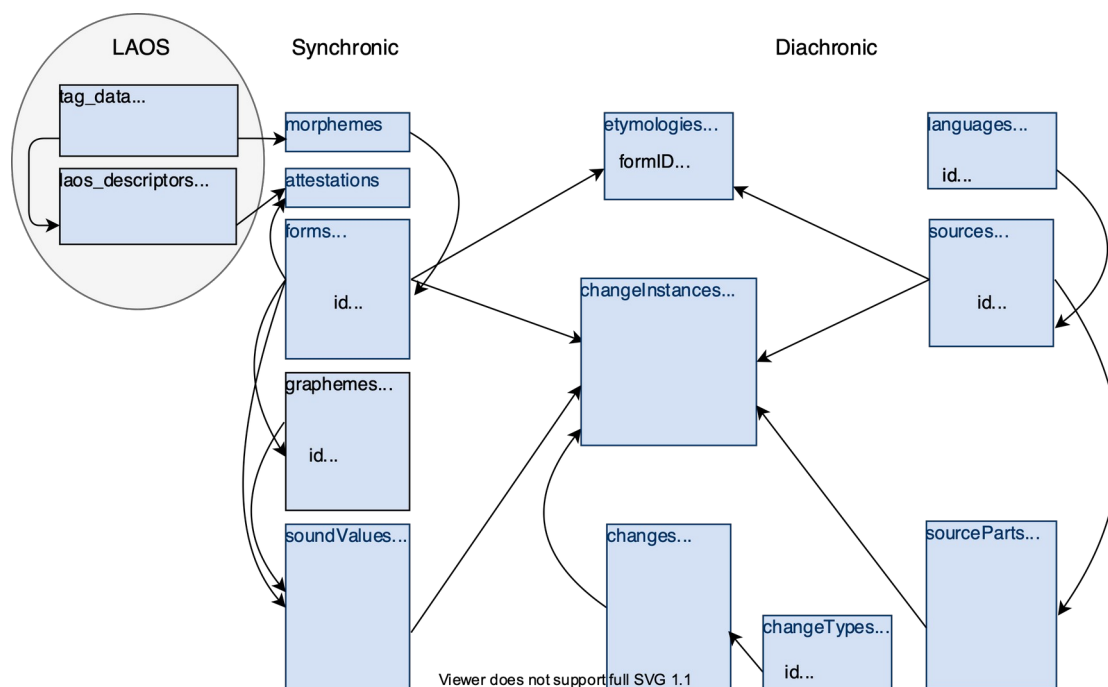

Figure 5. Schematic representation of the FITS corpus

Data entry was conducted via a series of bespoke tools allowing for parallel editing of a number of forms, keeping the structure outlined above. As a result, the search facility for users is also bespoke (Figure 6), allowing for searches across a wealth of different levels of analysis for the morpheme/word (present day English label, OSc spelling, OSc sound value, grammatical category) and metadata (text year, county of origin, LAOS number). It is also possible to search the database specifically from a diachronic standpoint, focusing on the Pre-Scots source for each morpheme, focusing on sound value, language tag or sound change. Most innovative, however, is the

---

[7] These are, by and large, root morphemes. We do not use the term *lemma*, as LAOS does, since in traditional lexicographic practice this refers to a 'base form' within a particular grammatical category (part of speech. Here we refer to the root form stripped of affixes and with no inherent link to a particular grammatical category, so the root *mother-* may surface in the adverbial *unmotherly*, the verb *mothered* or the noun *stepmothers*.

possibility of searching for sounds and spellings of individual segments — or series of these — across even more categories.

Figure 6: FITS search form: morpheme, metadata and sources (left); grapho-phonological context (right)

Results, which can be exported as CSV files, are given in table form with links to individual attestations, changes, source forms and texts, as well as token counts (Figure 7). Each morpheme headword is also hyperlinked to the relevant OED and *Dictionary of the Scots Language* (DSL) entry, in order to provide definitions, disambiguate and highlight contemporary reflexes.

Results can also be analysed temporally and geographically, as in Figure 8, showing the locations of <y> spellings for OSc [ø:] in the FITS mapping tool. These spellings occur in only 5 tokens of *doom* and *other* in Fife and Angus between 1459 and 1469. The pattern, therefore, appears to be the idiosyncrasy of a single scribe.

Form	Sound	Morpheme	Grammar	Texts Text ID–Year–County–Tokens
doume"	dø:m-e	doom OED DSL	n	1666–1447–ABD–1
dym~	dø:m-horizStroke	doom OED DSL	n	1275–1467–FIF–1
yy(er)is	ø:ð[][]-#infl	other OED DSL	npl	335–1459–AGS–1
ȝoull	jø:l	yule OED DSL	aj	1849–1495–FIF–1
uy(er)wayß	ø:ð[][]-±deriv	other OED DSL	av	453–1491–STG–1
vd(er)	ø:ð/d[][]	other OED DSL	aj	1277–1468–FIF–2

Total tokens	Source	Changes	Segments …
1	do:m (ONhb)	NF	**do:m** 2- 〈ou〉 -[ø:] (N)<NF<[o:] (N)
4	do:m (ONhb)	NF, PSE	**do:m** 2- 〈y〉 -[ø:] (N)<NF,PSE<[o:] (N)
1	o:ðer (ONhb)	NF	**o:ðer** 1- 〈y〉 -[ø:] (N)<NF<[o:] (N)
1	je:ol (OE)	EOM, NF	**je:ol** 2- 〈ou〉 -[ø:] (N)<NF,EOM<[e:o] (N)
1	o:ðer (ONhb)	NF	**o:ðer** 1- 〈u〉 -[ø:] (N)<NF<[o:] (N)
4	o:ðer (ONhb)	NF	**o:ðer** 1- 〈v〉 -[ø:] (N)<NF<[o:] (N)

Figure 7: Partial FITS results for a context-free search of [ø:], split and stacked for viewing convenience
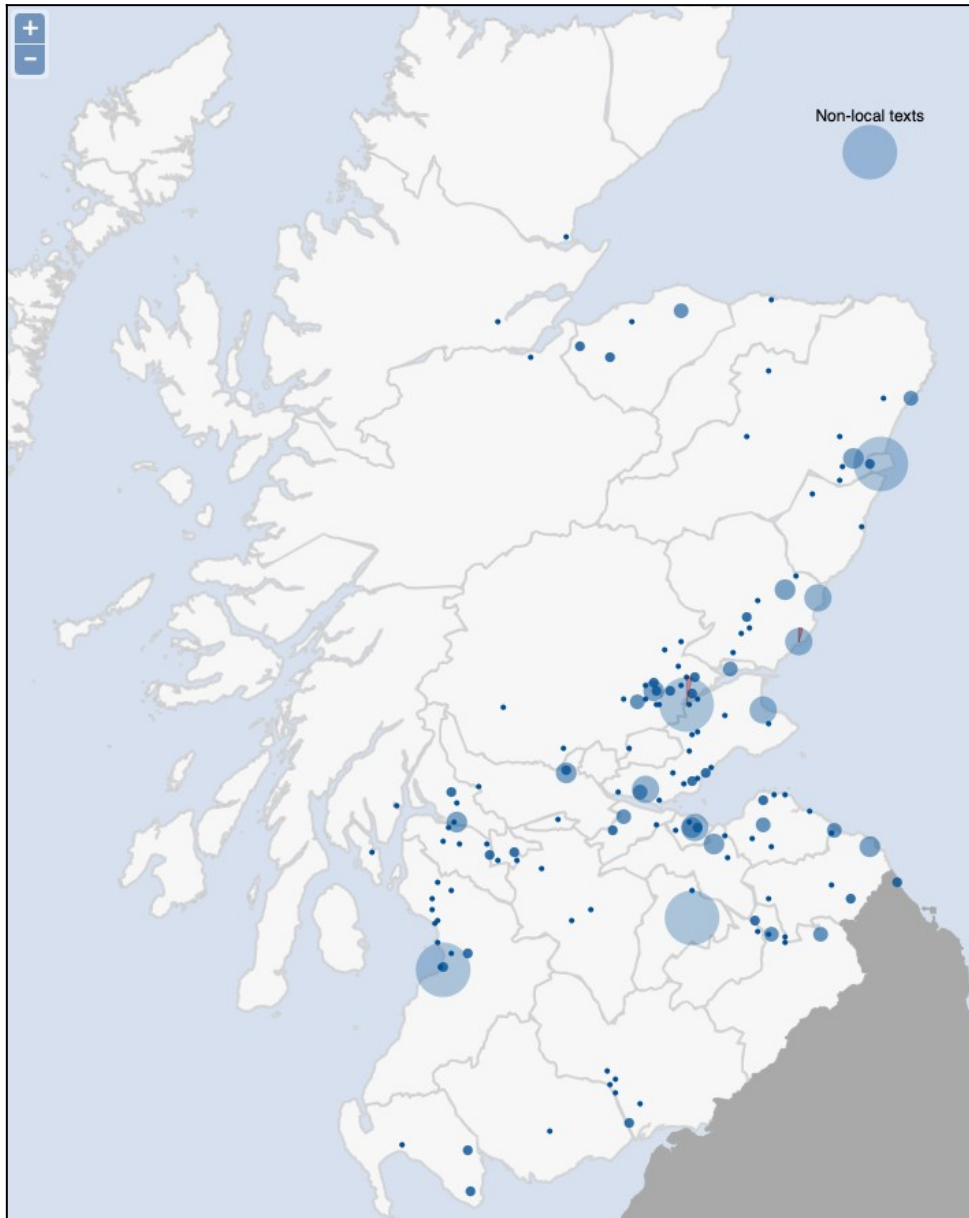
Figure 8: FITS-generated map showing the proportion of texts containing the target search (<y> spellings of OSc [ø:]) result by location. Red=match; blue=no-match

An additional feature of the FITS corpus already discussed in Kopaczyk *et al*. (2018: 264-5) is the grapho-phonological visualisation tool we refer to as *Medusa*. The tool allows us to represent graphemic (and sound) substitution sets dynamically, such as in Figure 9, which shows the array of spellings for the OSc segment [ø:] across the FITS corpus.

Figure 9: a graphemic substitution set for the segment [ø:]

4. Conclusions

As with any tagged corpus, FITS is not a neutral representation of data, but entails a substantial amount of explicit analysis of the available materials. Our intention as compilers has been to keep these analyses transparent throughout, both by explicitly stating etymological sources and changes, as well as keeping the dipl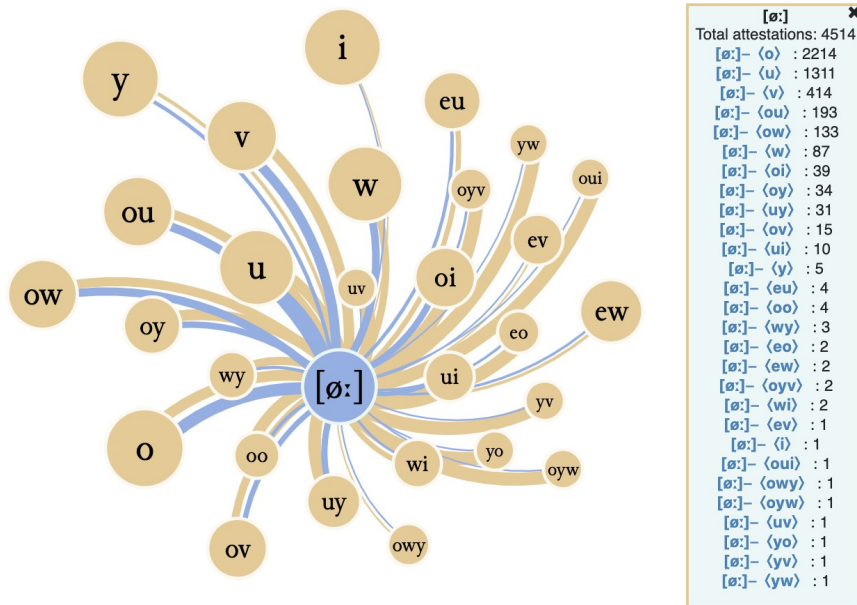omatically-transcribed spellings as the bedrock of our analyses, which will ultimately be open to users' own reinterpretations.

The FITS tools – which are freely accessible online – will enable end-users to interact dynamically with the OSc data and our analyses by:
- Investigating specific segments in terms of their sound values and spellings across multiple phonotactic, graphotactic and grammatical environments;
- Establishing temporal and spatial distributions for sounds and spelling across the corpus, using the textual metadata;
- Tracing the etymological sources of (Older) Scots morphemes and uncovering their specific phonological developments;
- Exploring proposed analyses for the emergence of the Older Scots forms via a Corpus of Changes linked to individual segments;
- Establishing connections with present-day English and Scots forms via links in the DSL and OED;
- Undertaking further quantitative and qualitative research into the data by downloading as CSV files or accessing the full, transcribed texts.

Whilst the FITS tools were developed specifically for the analysis of OSc, and are ultimately bespoke, they represent a proof of concept for analysing grapho-phonological relations and sound change via digital means more broadly. Indeed, some aspects of the project are already being applied to other languages, such as the use of grapho-phonological parsing and *Medusa* visualisations in the *Corpus of Historical Mapudungun* (CHM), and the application of graphemic variation analysis tools in the *LAEME Spelling Database Project* (Vaňková, 2021).

The development of FITS represents a substantial step forward in the application of corpus methods to historical phonological materials. Not only does this resource allow us to access a grapho-phonologically parsed body of pre-standard written material, it also presents us with a diachronic picture of these sound-spelling mappings. On the one hand, the individual linking of attested forms to etymological sources via sound changes is a way of supporting the synchronic analysis, but on the other, it also represents a robust proposal for a digitally-based history of the sound system of OSc, helping us uncover the distinct *pedigree* of this northernmost West Germanic variety.

References:

Aitken, A.J. and C. Macafee. 2002. *The Older Scots Vowels: A History of the Stressed Vowels of Older Scots from the Beginnings to the Eighteenth Century*. Edinburgh: Scottish Text Society.

Alcorn, R., B. Molineaux, J. Kopaczyk, V. Karaiskos, B. Los and W. Maguire. 2017. The emergence of Scots: Clues from Germanic *a reflexes. In *Before the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting, Ayr* 2015. Ed by J. Cruickshank and R. Mc. Millar. FRLSU, Aberdeen. 1–32.

Barrow, G. W. S. 2003. *The kingdom of the Scots: Government, church and society from the eleventh to the fourteenth century*. Edinburgh: Edinburgh University Press.

CHM: *Corpus of Historical Mapudungun (Version 1.0)*. Compiled by B. Molineaux, B. and V. Karaiskos. © The University of Edinburgh, http://www.amc-resources.lel.ed.ac.uk/CHM/.

CoNE: *A Corpus of Narrative Etymologies from Proto-Old English to Early Middle English and accompanying Corpus of Changes*. 2013. Compiled by R. Lass, M. Laing, R. Alcorn and K. Williamson. The University of Edinburgh. http://www.lel.ed.ac.uk/ihd/CoNE/CoNE.html.

DOST: A Dictionary of the Older Scottish Tongue 2004. Available online through the *Dictionaries of the Scots Language*. Scottish Language Dictionaries Ltd. <https://www.dsl.ac.uk >

DOEWC: *Dictionary of Old English Web Corpus* 2009. A. diPaolo Healey (ed.) Toronto: University of Toronto. <https://tapor.library.utoronto.ca/doecorpus/>

DSL: *Dictionary of the Scots Language*. 2004. Scottish Language Dictionaries Ltd. <https://www.dsl.ac.uk >

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

ECEP: *Eighteenth-Century English Phonology Database*. 2015. Compiled by J. C. Beal, N. Yáñez-Bouza, R. Sen, and C. Wallis. The Digital Humanities Institute, University of Sheffield. http://www.dhi.ac.uk/ecep.

FITS: *From Inglis to Scots: A Corpus of Grapho-phonological Correspondences (1380-1500) with Associated Corpus of Changes*. 2021– Compiled by R. Alcorn, A. Honkapohja, V. Karaiskos, J. Kopaczyk, B. Los, W. Maguire and B. Molineaux, © The University of Edinburgh. http://www.amc.lel.ed.ac.uk/fits/

Hogg, R. 2006. Old English Dialectology. In *The Handbook of the History of English,* A. van Kemenade and B. Los (eds.), Oxford: Wiley-Blackwell, 395–416.

Johnston, P. 1997a. Older Scots phonology and its regional variation. In *The Edinburgh history of the Scots language,* C. Jones (ed.), 47–111. Edinburgh: Edinburgh University Press. 47-111.

Johnston, P. 1997b. Regional variation. In C. Jones (ed.), *The Edinburgh history of the Scots language*, Edinburgh: Edinburgh University Press. 443–513.

Kopaczyk, J, B. Molineaux, V. Karaiskos, R. Alcorn, B. Los. and Maguire, W. 2018 Towards a grapho-phonologically parsed corpus of medieval Scots: Database design and technical solutions. *Corpora* 13(2):255–269.

Kries, S. 2007. Scottish-Scandinavian relations in the Middle Ages. The linguistic evidence. *NOWELE* 50/51: 109-130.

LAOS: *A Linguistic Atlas of Older Scots, Phase 1: 1380–1500*. Compiled by K. Williamson. 2013 [http://www.lel.ed.ac.uk/ihd/laos1/laos1.html] © The University of Edinburgh.

Macafee, C. and Aitken A.J. A history of Scots to 1700, in *A Dictionary of the Older Scottish Tongue.* 2002. vol. XII, xxix-clvii. Online: https://dsl.ac.uk/about-scots/history-of-scots/orthography/

Macafee, C. and A. Anderson. 1997. 'A random sample of Older Scots lexis'. *Transactions of the Philological Society* 95(2): 247-278.

McClure, J. D. 1981. Scottis, Inglis, Suddroun: Language labels and language attitudes. In Lyall and Riddy (eds.), *Proceedings of the Third International Conference on Scottish Language and Literature*. Stirling and Glasgow Universities, pp. 52–69.

Molineaux, B., Kopaczyk, J., Maguire, W., Alcorn, R., Karaiskos, V. and Los, B. 2019. Early spelling evidence for Scots L-vocalisation: a corpus-based approach. In *Historical Dialectology in the Digital Age.* Eds. R. Alcorn, B. Los, J. Kopaczyk and B. Molineaux, Edinburgh: Edinburgh University Press, pp. 61–87.

Molineaux, B., Kopaczyk, J., Maguire, W., Alcorn, R., Karaiskos, V. and Los, B. 2021. Phonotactics, graphotactics and contrast: The history of Scots dental fricative spellings. *English Language and Linguistics 25.1: 91–119*.

Molineaux B., W. Maguire, V. Karaiskos, R. Alcorn, J. Kopaczyk and B. Los. 2020. Visualising pre-standard spelling practice: The interchange of <ch(t)> and <th(t)> in Older Scots. *Journal of Data Mining and Digital Humanities, Special Issue on Visualisations in Historical Linguistics*, pp.1-11.

Murison, D. 1971. 'The Dutch element in the vocabulary of Scots'. In Aitken, A.J., A. McIntosh and H. Pálsson (eds.), *Edinburgh Studies in English and Scots*. London: Longmans. 159–176.

OED: *The Oxford English Dictionary Online.* 2021. http://www.oed.com. Oxford: Oxford University Press.

Proudfoot, E. and C. Aliaga-Kelly 1996. Towards an interpretation of anomalous finds and place-names of Anglian origin in North and East Scotland. *Anglo-Saxon Studies in Archaeology and History* 9: 1-14.

Samuels, M. L. 1985. The great Scandinavian belt. In *Papers from the 4th International Conference on English Historical Linguistics, Amsterdam, April 10–13, 1985*. Eds. R. Eaton, O.Fischer, W.F. Koopman and F. van der Leek. Amsterdam: John Benjamins. 269-82.

Vaňková, M. 2021. *LAEME Spelling Database Project*. Charles University, Prague. http://laeme-spelling.silent3.ff.cuni.cz/#/