



HAL
open science

Visualising pre-standard spelling practice: Understanding the interchange of ⟨ch(t)⟩ and ⟨th(t)⟩ in Older Scots

Benjamin Molineaux, Warren Maguire, Vasilios Karaiskos, Rhona Alcorn,
Joanna Kopaczyk, Bettelou Los

► To cite this version:

Benjamin Molineaux, Warren Maguire, Vasilios Karaiskos, Rhona Alcorn, Joanna Kopaczyk, et al..
Visualising pre-standard spelling practice: Understanding the interchange of ⟨ch(t)⟩ and ⟨th(t)⟩ in
Older Scots. *Journal of Data Mining and Digital Humanities*, 2020, Special Issue on Visualisations in
Historical Linguistics, Special issue on Visualisations in Historical Linguistics, pp.1-11. 10.46298/jd-
mdh.5568 . hal-02153662v2

HAL Id: hal-02153662

<https://hal.science/hal-02153662v2>

Submitted on 11 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visualising pre-standard spelling practice:

The interchange of <ch(t)> and <th(t)> in Older Scots

Benjamin Molineaux^{1*}, Warren Maguire¹, Vasilios Karaiskos¹, Rhona Alcorn^{1,2},
Joanna Kopaczyk³ and Bettelou Los¹

1 The University of Edinburgh, UK

2 Scottish Language Dictionaries, UK

3 The University of Glasgow, UK

*Corresponding author: Benjamin Molineaux benjamin.molineaux@ed.ac.uk

Abstract

Alphabetic spelling systems rarely display perfectly consistent one-to-one relationships between graphic marks and speech sounds. This is particularly true for languages without a standard written form. Nevertheless, such non-standard spelling systems are far from being anarchic, as they take on a conventional structure resulting from shared communities and histories of practice. Elucidating said structure can be a substantial challenge for researchers presented with textual evidence alone, since attested variation may represent differences in sound structure as well as differences in the grapho-phonological mapping itself. In order to tease apart these factors, we present a tool — *Medusa* — that allows users to create visual representations of the relationship between sounds and spellings (sound substitution sets and spelling substitution sets). Our case study for the tool deals with a longstanding issue in the historical record of mediaeval Scots, where word-final <cht>, <ch>, <tht> and <th> appear to be interchangeable, despite representing reflexes of distinct pre-Scots sounds: [x], [xt] and [θ]. Focusing on the documentary record in the *Linguistic Atlas of Older Scots* ([LAOS, 2013]), our exploration surveys key graphemic categories, mapping their lexical distributions and taking us through evidence from etymology, phonological typology, palaeography and historical orthography. The result is a novel reconstruction of the underlying sound values for each one of the target items in the record, alongside a series of sound and spelling changes that account for the data.

keywords

non-standard spelling; grapho-phonology; Older Scots; visualisation, sound change

INTRODUCTION

Throughout the mediaeval period, vernacular languages such as Older Scots (OSc) and Middle English (ME) lacked an orthographic standard. This meant there was no prescriptively correct way of spelling any given word, allowing scribes to draw from a range of orthographic conventions, mainly of Old English and (Anglo-)French origin to ‘sound out’ their own speech. The result was a variety of ways for spelling individual words, based on a multiplicity of ways of representing individual sounds. This variation depended partly on the scribe’s pronunciation, but also on lexical context (what word the sound came up in, whether the sound was at the start or end of a word, what the preceding letter was, etc.), and writing tradition (local, temporal, genre- or language-based) (for seminal work on scribal spelling systems, see [McIntosh, 1956], and [Samuels, 1963]). An example of this heterogeneity of spellings is given in Figure 1, representing a sample of the different forms for the word YEAR in the *Linguistic Atlas of Older Scots* corpus ([LAOS, 2013]) which brings together over 1,200 Older Scots legal texts from the period between 1380 and 1500.

yhere, zer(e), zhere, iere, yeer, yer, yhe, yheere, yheir, yheir(is), yheir(e), yhere, yhere(is), phere, zeere, zeir, zeir(e)-, zere, zerr, zer, zer(is), zeyr(e), zher, z^{er}(is) ...

Figure 1: a sample of the 118 spelling types for the word YEAR, out of 2,243 tokens across 857 texts in LAOS (letters in parentheses are expanded abbreviation symbols)

Ultimately, each sound (or *phone*) in medieval spelling could have several spellings – a spelling substitution set – and, vice-versa, each spelling unit (or *grapheme*) could be used to represent multiple phones – a sound substitution set (cf. [Laing, 1999], [Laing and Lass, 2003], [Kopaczyk et al., 2018]). Figure 2 gives example visualisations of such sets, based on the data in the *From Inglis To Scots Corpus* ([FITS, forthcoming]), a unique resource for exploring non-standard spelling practices. In what follows, we give an overview of the challenges and advantages of creating and using such visualisations. The FITS corpus contains the spellings of more than 100,000 word tokens from the LAOS corpus. In order to fully explicate these tools, we focus on the reconstruction of a particularly difficult set of overlapping sound and spelling substitution sets, those involving the spellings <ch>, <cht>, <th> and <tht>, among others.¹

Visualisations such as those in Figure 2 are intended to help researchers understand the diversity of sound-spelling correlations in a corpus, as well as the relative frequency of each phone (size of the blue bubble), grapheme (size of the mustard bubble) and the links between them (thickness of a line represents a proportion of the total tokens). Such information – in this case for OSc – helps guide the reconstruction of the spoken language of particular periods and locations and trace linguistic change over time and space. Methodologically, these substitution sets are made possible by the careful construction of a grapho-phonologically parsed corpus – which matches reconstructed phones to graphemes (see [Kopaczyk et al., 2018] for details) – linked to a custom-built visualisation tool. The latter is a weighted network graph built with the D3 Javascript library, and which we have christened *Medusa* (cf. [Alcorn et al., forthcoming]).

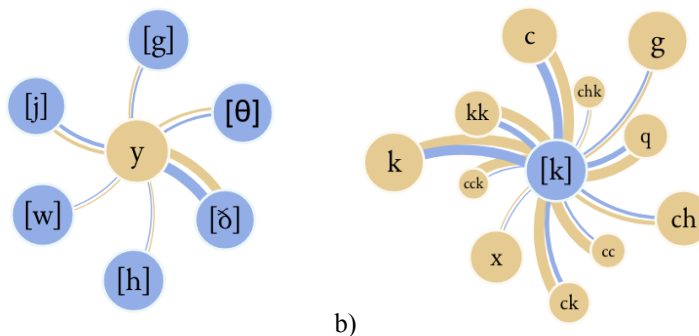


Figure 2: FITS Corpus substitution sets for (a) the consonantal sounds associated with grapheme <y> and (b) the spellings representing the phone [k].

I BUILDING HISTORICAL SUBSTITUTION SETS AND RECONSTRUCTING CHANGE

In order to reconstruct the sound values behind spelling forms, we must triangulate from a number of factors. Spelling is one, of course, but the etymological source of a given word and

¹ Throughout the article, we use the standard linguistic convention of placing graphemes in angled brackets < > and phones in square brackets []. For the latter, we use the symbols of the [International Phonetic Association](http://www.internationalphoneticassociation.org). They are not intended to represent allophonic variation but “a reasonable transcriptional response” ([Lass and Laing, 2013: Section 2.4.2]) to the sound value aimed at by the scribe. Present-day English equivalents of words are given in small caps.

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 in question. For every instance of a grapheme, therefore, FITS provides a triad made up of (i) an OSc grapheme, (ii) an OSc phone and (iii) a Pre-Scots, etymological phone (usually from Old English, but see [Alcorn et al., 2017]), as in Figure 3. Where there is a discrepancy between the Pre-Scots and the OSc phones, the relevant linguistic development is reconstructed. In the example in Figure 3, the change from [i] to [ɪ] is characterised as an instance of Short Vowel Lowering (SVL), following [Aitken and Macafee, 2002:10-11]. Ultimately, FITS can be seen as a corpus of such triads alongside a corpus of linguistic changes.

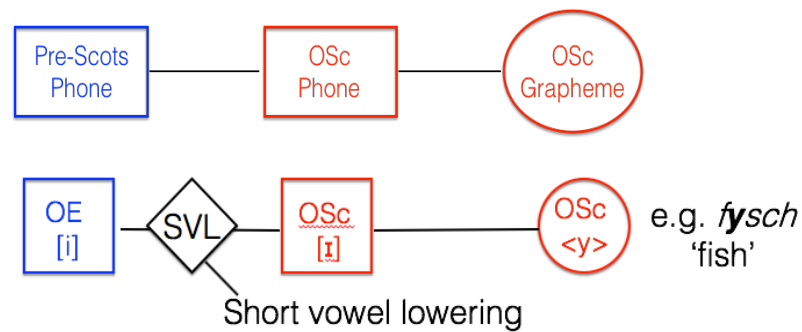


Figure 3: A schematic representation of source-sound-spelling triads in FITS (above) and a single triad for the grapheme <y> in the spelling *fyſch* for OSc FISH attested in the corpus (below)

II CASE STUDY: GRAPHEMIC OVERLAP OF <CH(T)> AND <TH(T)>

A famously difficult pattern of sound-spelling mappings in the early history of Scots – and one that needed a solution for the grapho-phonological parsing of the FITS Corpus – is that of the ends of words such as *burgh*, *thought* or *cloth*. In such words, OSc scribes used an overlapping range of spellings for the reflexes² of Pre-Scots [x], [xt] and [θ]. Indeed, according to [Johnston, 1997a: 101], the Older Scots spellings “<ch, cht, th, tht> appear interchangeable, leading some authorities to conclude that they are just graphical variants”. In other words, here we have a number of cases where etymologically distinct sound categories (Pre-Scots [x], [xt] and [θ]) have at least partially merged spellings (<ch, cht, th, tht>). This prompts questions such as whether the phonic substance of these spellings has also merged and what changes to the sound and spelling systems need to be postulated in order to account for the spelling overlap.

The pattern is complex in that it entails three seemingly unrelated linguistic changes:

- a. Final <t> being lost in many OSc forms derived from Pre-Scots [xt] words:³
thoch THOUGHT, *auch* EIGHT, *boch* BOUGHT, *frach* FREIGHT, *rich* RIGHT, *weich* WEIGHT, *myth* MIGHT
- b. A non-etymological final <t> often appearing in OSc forms derived from Pre-Scots [x] and [θ]-final words:
burght BURGH, *throucht* THROUGH, *laucht* LAW, *yotht* THOUGH, *southt* SOUTH, *truetht* TRUTH, *clacht* CLOTH, *vortht* WORTH, *aitht* OATH
- c. Final <ch(t)>~<th(t)>(~<gh(t)>) alternating with each other in OSc spelling:

² By ‘reflexes’ we mean the sounds found in Older Scots in the same positions and for the same words as in their Pre-Scots etymological sources (Old English, Old Norse or otherwise).

³ All examples in a), b) and c) taken from the FITS corpus.

monecht~moneth MONTH, *norcht~north* NORTH, *clacht~clath* CLOTH,
furcht~furht~furcht FORTH, *acht~ath* OATH, *tolbuch~tolbuth* TOLLBOOTH,
lench~lenth LENGTH

A glimpse of the complexity of these overlaps is evident from the Medusa-generated OSc spelling substitution sets for word-final [θ] and [x] (from Pre-Scots [θ] and [x]), in Figure 4. Recall that the thickness of the blue line joining sounds to spellings gives an indication of the proportion of phone tokens represented by the relevant grapheme, so that the main variants here are clearly <ch>, <th>, <cht> and <tht> for both etymological categories. The superscript <^t> (represented as <^t>) is an important variant for [θ]-words, and a minor one for [x]-words. Finally, [x] alone is represented by <gh>, for which it is a major variant.

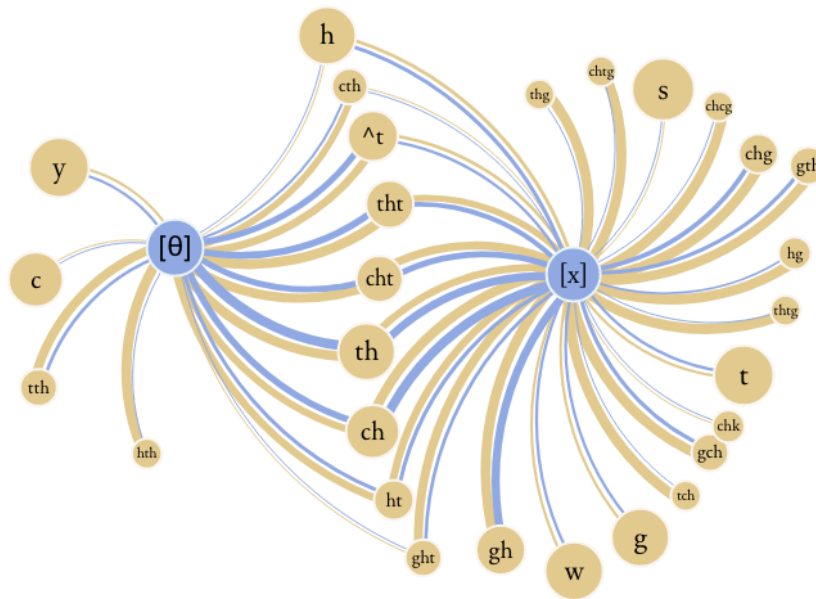


Figure 4: Overlapping OSc spelling substitution sets for word-final [θ] and [x] in the FITS Corpus

A quantitative view of the number of spelling tokens for each of the relevant Pre-Scots sound categories – [θ], [x] and [xt] ([θt] did not exist) – can also be obtained through the corpus tools, such as those underlying Figure 5. A careful examination of these 5,127 tokens reveals that, although most spellings are found in all historical sound categories, their distribution is quite different. In what follows, we look at this diversity of spellings to postulate underlying OSc phonic structures and a scenario for their development from the Pre-Scots sources.

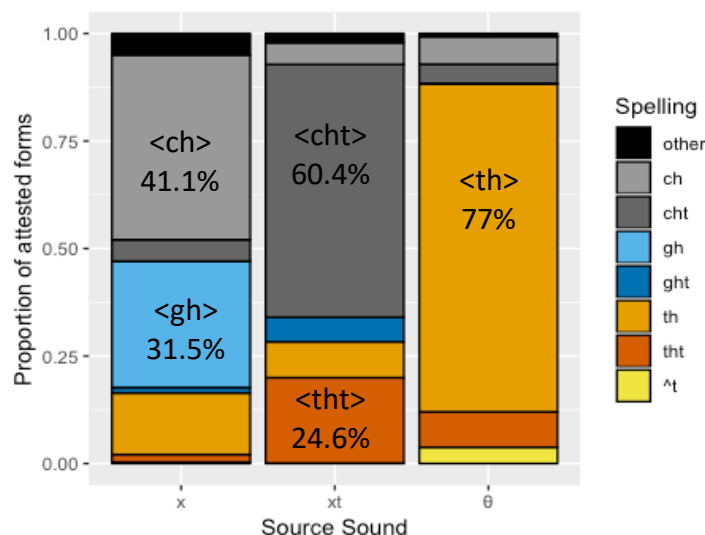


Figure 5: Proportion of OSc attested spelling tokens derived from the Pre-Scots [x], [xt] and [θ] categories

2.1 Expected spellings

The pattern in Figure 5 shows the three Pre-Scots sets having, for the most part, their ‘expected’ spellings in FITS. Etymological (i.e. Pre-Scots) [x] is mostly spelled <ch> and <gh> in OSc, which (alongside <h>) are the main Old English representations of the velar fricative (<gh> being far rarer and mostly northern, see [Hogg, 1991: 286]). The vast majority of [xt] forms are spelled as <cht> or <ght> following the same expected pattern for [x]. As for etymological [θ], the main variant, <th>, is also expected. Under (Anglo-)Norman pressure, a dispreference developed towards the native English <ð> spelling of word-final [θ]. As a result, the <th> digraph was brought into Northern Middle English and early Scots from Anglo-Latin, where it had long been used to spell dental fricatives, first in Greek <θ>-words, and then in Old English names (cf. [Benskin, 1977] and [Molineaux et al., 2020]).⁴

For the expected spellings, we can safely assume that the underlying sounds are also unchanged from their etymological values, e.g. that <cht> represented [xt] in OSc, as it did in Old English. Nonetheless, a substantial minority of tokens have spellings whose sound values are not so evident. Among these are the <th> and <tht> spellings for etymological [x] (17.3%); the <th> and <tht> spellings for [xt] (32.2%), and <ch> and <cht> for [θ] (10.4%). How should these be explained? Are these simply spelling innovations and, if so, what was their motivation? Or do they signify that the original sound values had subsequently changed? As we assume scribes were not randomly choosing their spelling repertoires and were “capable of sophisticated and subtle linguistic analysis” ([Laing and Lass, 2003: 258]), we postulate that these choices encode worthwhile information, either at the level of the orthography or the sound system. In the following subsections we attempt to elucidate the likely reasons behind such choices.

2.2 The rise of <gh> and the optimisation of graphemic contrast

While both <ch> and <gh> are expected spellings for OSc [x], as in night being spelled <nicht> or <night>, and both were available in Old English scribal practices, <gh> was quite rare, so it

⁴ OSc <^t> is attested 128 times in words which had Pre-Scots [θ] and only four times in forms derived from Pre-Scots /x/ words. In general, it may be straightforwardly considered an abbreviation of <th>.

is surprising how frequent it becomes in the FITS data.⁵ Given the 120-year span of our corpus, we are able to look at the development of spellings for this category, to try and identify a diachronic trend. Figure 6 shows the proportion of graphemic representations for etymological [x] over the decades of the FITS Corpus.⁶ While our earliest data is rather sparse, the overall trend in the later data indicates a reduction in the use of <ch> and an increase in the use of <gh> to represent the velar fricative. While both these options were available, we suggest that, from the Old English period, the rise of <gh> represents an optimisation of spelling contrast. The visualisation in Figure 7 provides the motivation for this: besides representing the velar and dental fricatives, <ch> could also stand for a palato-alveolar affricate, [tʃ] (see [Bann and Corbett, 2015: 23]). On the other hand, <gh> only represents [x] and its variant before front vowels, [ç]. Ultimately, since <ch>=[x/ç] had more sound representations to compete with than <gh>=[x/ç], <gh> was, among the two, the optimal candidate for this role.

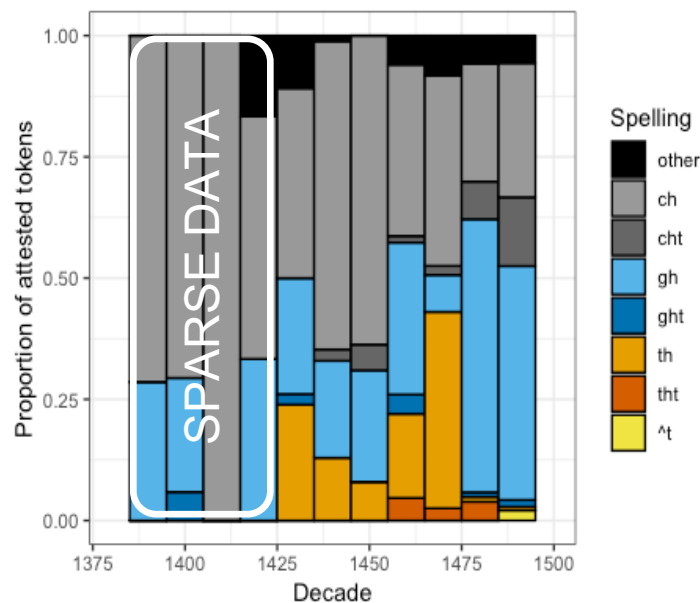
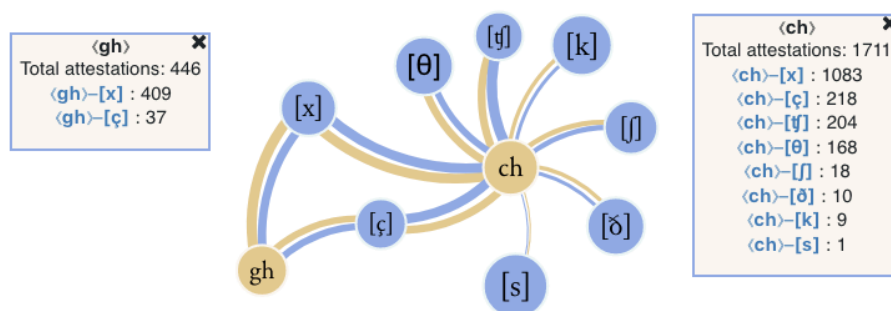


Figure 6: Proportion of OSc graphemes for tokens of etymological [x] by decade in the FITS corpus



⁵ An important pattern that one of the reviewers rightly highlights is that of the strong lexical skew among <gh> spellings, which occur most often in the word BURGH (63%). Given that LAOS texts are predominantly administrative documents from early urban centres, this is an extremely frequent item. It is therefore a relevant question whether this pattern obtains outside of our target textual genre.

⁶ While our focus here is primarily on the spelling alternations in the corpus, there is also a significant phonological process that affects the frequency of etymological [x] spellings in OSc. This is the process of vocalisation of [x], usually via an intermediate process of weakening (see [CoNE, 2013: *Gamma Weakening* –GW, and *Coda Vocalisation* – CV]). The result, as a reviewer points out, is that the phonological outcome of [x] is often a vocalic realisation (cf. *borrow* BURGH). Our analysis, focusing on the consonantal spellings, leaves these forms aside.

Figure 7: Overlapping sound substitution sets for <gh> and <ch> across the FITS corpus

2.3 Final <t> loss and hypercorrective <t> insertion

The loss of final <t> in etymological [xt] forms and the simultaneous appearance of unetymological <t> in reflexes of [x]- and [θ]-final morphemes (see Figure 5) has been noted in the literature on Older Scots phonology. The first process has been linked to the more general, though sporadic and non-localised phenomenon of /t/-deletion in the morpheme-final clusters [xt, st, ft, pt, kt] as in *deiching* (<*dichting*) preparing, *cassen* (<*casten*) cast (past participle), *giff* gift, *excep* except, *contrak* contract ([Johnston 1997a: 101], [CoNE, 2013: *Final Coronal Deletion* – FCD]). This would, of course, account for Pre-Scots [xt] becoming Older Scots [x], spelled <ch> (cf. [Meurman-Solin, 1997]). In present day Scots dialects, the outcome of this change has been mostly ‘restored’ to [xt] (cf. [Romaine, 1984]). however, there are exceptions. Indeed, [Macafee and Aitken, 2002: §5.2, fn. 87] note that the reflexes of [xt]>[x] can still be heard “in Lanarkshire speech at the present time”.

The counterpart of this [t]-deletion process is what [Kniezsa 1997: 43] calls ‘intrusive <t>’ and [Johnston 1997a: 101] calls “excrecent /t/”. What is entailed by these terms is the addition of either an unetymological sound, i.e. [t], or letter, i.e. <t>, after a morpheme-final vowel+consonant sequence, as in burgh spelled <burcht>. At least since [Murray 1873: 128], this has been attributed to hypercorrection, that is, to the speakers’ uncertainty as to the lexical incidence of final [t], following its sporadic loss (see also [Meurman-Solin, 1997], [Romaine, 1984], [CoNE, 2013: *Final Consonant Excrescence* – FCE]). Whether the hypercorrection is phonic, as claimed by [Johnston, 1997a: 101] and [Romaine, 1984], or purely orthographic (a so-called *backspelling*), the alternation between <ch> and <cht> for etymological [x] and [xt] (as well as between <th> and <tth>, as we shall see) can be seen as a product of [t]-deletion and consequent hypercorrection.

2.4 [xt]-dentalisation and the <cht>~<th> overlap

An additional change which shows up in the literature on Older Scots is the process of [xt] dentalisation, which is invoked as an explanation for OSc <th> and <tth> spellings for Pre-Scots [xt] in forms such as *dother* DAUGHTER and *vortht* WORTH. Evidence for the reality of this as a sound change comes from the fact that its outcome is still visible in some North-Eastern dialects of Scots today ([Macafee and Aitken 2002: §5.2 fn.87]). According to [Johnston, 1997b: 505] (see also [Grant 1931: xxxv], [Dieth, 1932: 113], [CoNE, 2013: *Transposition x to Theta* – TXT]), this process can be thought of as an instance of progressive place assimilation where the first element of the consonant cluster is articulated in the same place of the second element ([xt]>[θt]).⁷ This would explain OSc <tth> spellings for Pre-Scots [xt] as instances of ‘fronting’ to [θt]. In the cases where the sequence surfaces as <th>, the <t> element would be subsequently subject to [t]-deletion (cf. §2.3, above) and therefore represent an instance of [θ].⁸ In this account of [xt] dentalisation, the change of [x]> [θ] is only motivated for the etymological [xt] category, not for Pre-Scots [x]. However, [CoNE, 2013: *Transposition x to*

⁷ While here we use the symbol [t] for the second element of this cluster, both in Pre-Scots and OSc, it is likely that for much of this period the articulation was probably dental [t̪], thus explaining the assimilation of [x] (velar) to [θ] (dental).

⁸ If fronted [xt] variably retained <ch(t)> as a spelling, then it is possible that the spelling became licensed, in some cases and scribal systems, to represent [θt] or [θ], however below we argue for transcriptional confusion as the main source for such spellings.

Theta – TXT] claims that “less commonly” [x] > [θ] occurs without the following [t], word finally. While this seems to coincide with some of the spellings in FITS (such as FITS *bourth* BURGH), lack of phonic motivation and absence of present-day reflexes for such a change – at least in the Scots context – lead us to believe that this is merely a graphemic change, explained in §2.5, below.⁹

While it is likely that OSc <th> spellings corresponding to Pre-Scots [xt] retain a [t] in pronunciation, it is also possible that in some cases the <t> is retained in the spelling alone, as part of the more general ambiguity of this spelling (as per §2.3). In the case of <th> spellings derived from Pre-Scots [θ], final <t> is most likely only graphic, as [θt] is not a part of any Pre-Scots inventories, providing little ground for hypercorrective pronunciations. Furthermore, as [Murray, 1873:128] points out, <tht>- and <th>-final words are happily rhymed throughout the period and “the spellings *mouth, mouitht, zenyth, zenytht, with, witht*, are found promiscuously on the same page of early books and MSS”. Interestingly, [Meurman-Solin, 1997: 121] notes that <tht> is a “spelling practice” found predominantly in formal and legal texts associated with central administration during the fifteenth century and that it only spreads to more informal registers and the periphery beginning in the sixteenth century, disappearing altogether in the early seventeenth century. Taking all this evidence together, we consider OSc <tht>, where derived from Pre-Scots [θ], to represent unchanged OSc [θ].

2.5 Letter-shape overlap and <ch>~<th> transcription confusion

As for the use of <ch> in place of etymological [θ], as well as <th> for etymological [x], the explanation is likely not to do with the spelling system, but rather with letter shapes and their subsequent transcription by the compilers of editions. Indeed, it has been claimed that “[t]he interchange of <c>, <t> in this set is perhaps adequately explained by purely orthographic considerations, namely the confusability in secretary hand of <c> and <t>” (Macafee and Aitken, 2002: fn. 87). This is borne out in the manuscripts underlying FITS, where the two letter shapes are typically indistinguishable (Figure 8). However, such *figurae* or *allographs* (see [Benskin, 1997] and [Kopaczyk et al., 2018], respectively) are almost never transcribed as <c> in Pre-Scots [t] contexts by present-day scholars and editors working on OSc texts. Pre-Scots [k] is never given as <t> either, despite overlapping shapes. This is probably due to the fact that transcribers are aware that there are no changes of [t]>[k] or [k]>[t] proposed in the literature, thus eliminating these transcription choices in published forms of the MSS (cf. Figure 9). Likewise, etymological [tʃ] and word-initial [x] are almost never spelled <th> in transcriptions, since [tʃ]>[θ] and [x] > [θ] word-initially are not viable changes.

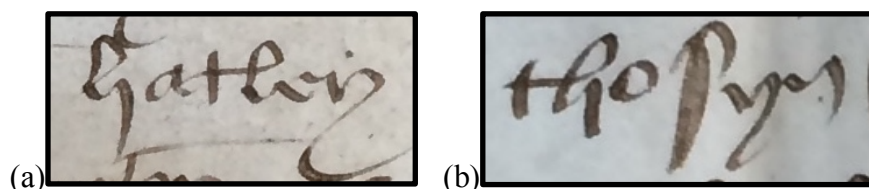


Figure 8: Images of the words transcribed as (a) *hatley* Hatley(surname) and (b) *chosyn* chosen, LAOS text N° 5 (1480 charter from the Laing Charters n° 15, box 1; Edinburgh University Library)

⁹ An additional suggestion made by a reviewer is that, in cases such as *burth* BURGH and *bourth* BORCH ‘surety’, [x] > [θ] could be the result of place assimilation to the preceding [r]. Lack of later survival of this [θ] and the unavailability of similar explanations for forms such as *pluth* PLOUGH and *mauth* MAUCH ‘son in law’ argue against the idea of and [r]-conditioned assimilation from velar to dental.

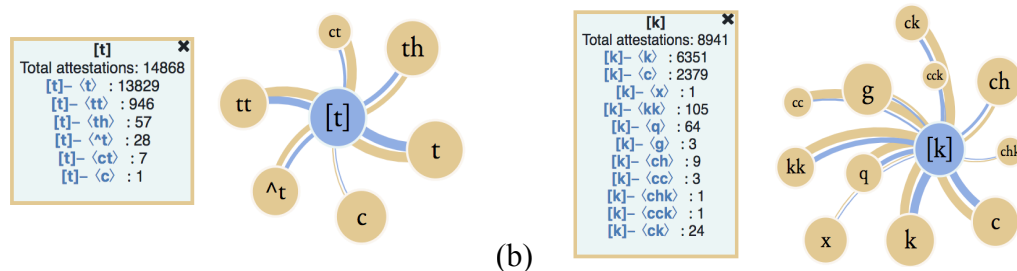


Figure 9: Graphemic substitution sets for (a) [t] and (b) [k] in the FITS Corpus.

On the other hand, the literature suggests that Pre-Scots [xt] *can* be realised as OSc [θ(t)]. This leads transcribers to carefully consider the near-identical initial shapes in the digraphs (or trigraphs), as <ch(t)> is taken to represent the un-fronted variant [x(t)] (Figure 10a) and <th(t)> the fronted one ([θ(t)], cf. Figure 10b). This leads to a shape-based appraisal of all such graphs (i.e. the ambiguous <c~t>) when followed by <h(t)> in non-initial position. More problematically, however, it appears that transcribers take this approach not only with etymological [xt] words, but also with etymological [θ] and [x] words. Consequently, where the relevant shape is deemed more <c>-like, etymological [θ] is transcribed as <ch(t)> (Figure 10c), despite no proposed process of [θ]>[x(t)]. We consider these cases to be miscategorised <ch(t)> spellings which should be merged in favour of <th(t)>. Likewise, in etymological [x] cases where the relevant *figura* is seen as <t>-like, it is erroneously transcribed as <th(t)> as we see in Figure 10d. Since we have rejected a word-final [x]>[θ] change, we assume that such cases simply represent a continuation of Pre-Scots [x] into OSc.

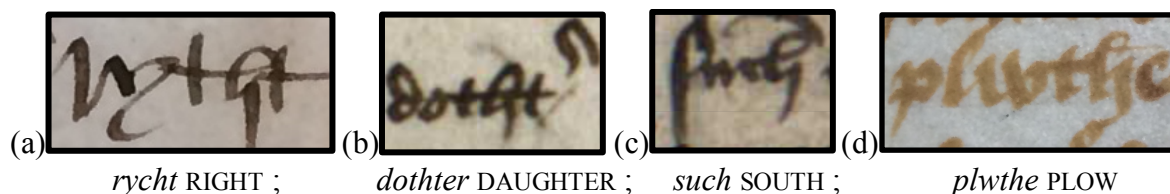


Figure 10: Images of the words transcribed as (a) *rycht* RIGHT in LAOS text N° 5 (see Fig. 8), (b) *doht(er)* DAUGHTER in LAOS text N° 1691 (1449/50 entry in the Aberdeen Council Register, p.76 [image]), (c) *such* SOUTH in LAOS text N° 1707 (1452 entry from the Aberdeen Council Register, p. 156 [image]) and (d) *plwthe* PLOW in LAOS text N° 341 (1459 entry from the Arbroath Abbey ‘Registrum Nigrum’, National Library of Scotland Adv. MS 34.4.3, f.80r.)

2.6 Accounting for the data: a new set of triads and changes

The apparently random mass of variant spellings in OSc word-forms containing Pre-Scots [xt], [x] and [θ] can be almost entirely explained by appeal to three well-known sound changes in the history of Scots: [t]-deletion (TD), [t]-excrescence (TE), and [xt]-dentalisation (‘place assimilation’ – PA), alongside some instances where transcribers have confused (near)identical <t> and <c> (‘transcription confusion’ – TC). Further spelling-only processes, such as spelling set optimisation (SSO, cf. the rise of <gh> in §2.2), abbreviation (ABBR) of <th> to <^t>, and final <t> backspelling (FTBS) complete the picture.

A close analysis of the data suggests that the OSc scribes followed clear principles mapping spellings to sounds, even if these led to overlapping spelling substitution sets. This consistency allows us to reconstruct the sound values for each one of the spellings given in the FITS Corpus with a large degree of confidence and insert them within triads such as those in Figure 3. Table

1 gives a detailed breakdown of reconstructed triads for our target categories, invoking relevant sound and spelling changes that link them together.¹⁰

Pre-Scots sound value	Sound change	OSc sound value	Spelling innovation	OSc spelling unit	FITS Example	% of category
[x]		[x]		<ch>	<i>through</i> THROUGH	41.1%
		[x]	SSO	<gh>	<i>burgh</i> BURGH	31.5%
		[x]	TC	<th>	<i>awth</i> EIGHT	15.6%
	TE [?]	[x]~[xt]	FTBS [?]	<cht>	<i>throught</i> THROUGH	4.5%
	TE [?]	[x]~[xt]	TC + FTBS [?]	<tht>	<i>yoht</i> THOUGH	1.8%
	TE [?]	[x]~[xt]	SSO + FTBS [?]	<ght>	<i>burght</i> BURGH	1.1%
[xt]		[xt]		<cht>	<i>acht</i> EIGHT	60.4%
	PA	[θt]		<tht>	<i>myht</i> MIGHT	19.7%
	PA+TD	[θ]		<th>	<i>dowthir</i> DAUGHTER	7.6%
		[xt]	SSO	<ght>	<i>knight</i> KNIGHT	5.9%
	TD	[x]		<ch>	<i>auch</i> OUGHT	4.5%
[θ]		[θ]		<th>	<i>baith</i> BOTH	77%
		[θ]	FTBS	<tht>	<i>aith</i> OATH	7.9%
		[θ]	TC	<ch>	<i>clach</i> CLOTH	6%
		[θ]	ABBR	<^t>	<i>fur^h</i> FORTH	4.3%
		[θ]	TC + FTBS	<cht>	<i>norcht</i> NORTH	4%

Table 1: Summary of proposed OSc phones by etymological category and spelling form in the FITS corpus, with relevant sound and spelling changes¹¹

III CONCLUSIONS:

The FITS Corpus is a unique resource for exploring the historical relationship between sounds and spellings. While it relies on traditional historical phonological practices to establish likely grapho-phonological links, these are facilitated and enhanced by visualisation tools. Among such tools, we have shown the advantages of Medusa-type graphemic and sound substitution sets (Figures 2, 4, 7, 9), as well as graphs representing the proportions of OSc spellings in a given category, either across the entire corpus (Figure 5) or diachronically (Figure 6).

While we have seen how grapho-phonologically parsed corpora, alongside key visualisation tools, can be used to tease apart overlapping sound and spelling patterns in Older Scots, their applicability is far broader. Ultimately, FITS provides a suite of tools for generating and revising likely scenarios for sound or spelling changes including:

- A means to visualise relationships between historical sounds and spellings
- A way of quantifying these relationships and link them to linguistic and extralinguistic factors
- A way to visualise the distribution of sounds and spellings across time and space

¹⁰ As shown in Figure 7, the FITS Corpus assumes that, adjacent to high and mid front vowels, Older Scots [x] surfaced as [ç] (in phonological terms, these two are allophones). This alternation is not represented in Table 1 for simplicity's sake, but we do indeed assume words like *nicht* MIGHT were articulated with a palatal fricative, [ç].

¹¹ Superscript question-marks in the table represent cases where there is likely alternation between phonologically excrecent [t] and backspelled <t> in the same spelling category, and where establishing a definitive distinction between the two is impossible.

We hope such resources will be used widely as applied to the FITS Corpus itself and that the methods – now tested for proof of concept – will in the future be harnessed for analyses of other periods and language families.

References

- Alcorn, R., Molineaux, B., Kopaczyk, J., Karaiskos, V., Los, B. and Maguire, W. The emergence of Scots: Clues from Germanic *a reflexes. In J. Cruickshank and R. Mc. Millar (eds.), *Before the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting, Ayr 2015*. Forum for Research on the Languages of Scotland and Ulster (Aberdeen). 2017:1–32.
- Aitken, A. J. and Macafee, C. *The Older Scots Vowels*. Scottish Text Society (Edinburgh), 2002.
- Bann, J. and Corbett, J. *Spelling Scots. The orthography of literary Scots, 1700-2000*. Edinburgh University Press (Edinburgh), 2015.
- Benskin, M. Local archives and Middle English dialects, *Journal of the Society of Archivists*. 1977; 5(8): 500-514.
- Benskin, M. Texts from an English township in late medieval Ireland, *Collegium Medievale*. 1997;1(2): 91–170.
- CoNE = Lass, R., Laing, M., Alcorn, R. and Williamson, K. (compilers). *A Corpus of Narrative Etymologies from Proto-Old English to Early Middle English and accompanying Corpus of Changes*. The University of Edinburgh (Edinburgh). 2013. <http://www.lel.ed.ac.uk/ihd/CoNE/CoNE.html>
- Dieth, E. *A grammar of the Buchan dialect*. Heffer (Cambridge), 1932.
- FITS = Alcorn, R., Honkapohja, A., Karaiskos, V., Kopaczyk, J., Los, B., Maguire, W. & Molineaux, B. (compilers). *From Inglis to Scots: A Corpus of Grapho-phonological Correspondences (1380-1500) with Associated Corpus of Changes*. The University of Edinburgh (Edinburgh). Forthcoming. <http://www.amc.lel.ed.ac.uk/fits/>
- Grant, W. Introduction, *The Scottish National Dictionary*, vol. 1, Scottish National Dictionary Association (Edinburgh). 1931.
- Hogg, R. M. *A Grammar of Old English*, Vol 1: Phonology, Oxford University Press (Oxford), 1992.
- Johnston, P. Older Scots phonology and its regional variation. In Charles Jones (ed.), *The Edinburgh history of the Scots language*, Edinburgh University Press (Edinburgh). 1997a: 47–111.
- Johnston, P. Regional variation. In C. Jones (ed.), *The Edinburgh history of the Scots language*, Edinburgh University Press (Edinburgh). 1997b: 433–513.
- Kniezsa, V. The origins of Scots orthography. In C. Jones (ed.), *The Edinburgh history of the Scots language*, Edinburgh University Press (Edinburgh). 1997: 24-46.
- Kopaczyk, J., Molineaux, B., Karaiskos, V., Alcorn, R., Los B. and Maguire, W. Towards a grapho-phonologically parsed corpus of medieval Scots: Database design and technical solutions. *Corpora* 2018;13(2):255-269.
- Laing, M. Confusion *wrs* confounded: Litteral Substitution Sets in Early Middle English writing systems, *Neuphilologische Mitteilungen*. 1999;100(3):251-270.
- Laing, M. and Lass, R. Tales of 1001 nists: The phonological implications of litteral substitution sets in some thirteenth-century South-West Midland texts, *English Language and Linguistics* 2003;7(2):257-278.
- LAOS = Williamson, K. (compiler). *A Linguistic Atlas of Older Scots, Phase 1: 1380–1500*. The University of Edinburgh (Edinburgh), 2008. <http://www.lel.ed.ac.uk/ihd/laos1/laos1.html>
- Lass, R. and Laing, M. Introduction, Chapter 2, in *A Linguistic Atlas of Early Middle English, 1150–1325*. (Version 3.2.) Compiled by M. Laing. University of Edinburgh (Edinburgh), 2013. <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>
- 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/>
- McIntosh, A. The analysis of written Middle English. *Transactions of the Philological Society*. 1956:26-55.
- Meurman-Solin, A. Differentiation and Standardisation in Early Scots. In C. Jones (ed.), *The Edinburgh history of the Scots language*. Edinburgh University Press (Edinburgh), 1997:3-23.
- Molineaux, B., Kopaczyk, J., Maguire, W., Alcorn, R., Karaiskos, V. and Los, B. Phonotactics, graphotactics and contrast: The history of Scots dental fricative spellings. *English Language and Linguistics (first view)*, 2020:1-29. doi:10.1017/S1360674319000479.
- Murray, J. *The dialect of the southern counties of Scotland*. The Philological Society (London), 1873.
- Romaine, S. The sociolinguistic history of /v/d deletion. *Folia Linguistica Historica*. 1984;5(2):221-255.
- Samuels, M. L. Some Applications of Middle English Dialectology. *English Studies*. 1963;44:81–94, reprint in *Middle English Dialectology: Essays on Some Principles and Problems*, ed. M. Laing (Aberdeen). 1989:64–80.