The FITS Corpus:
Tracing the origins of fifteenth-century Scots sounds and spellings

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

23 August, 2018

The FITS project (*From Inglis To Scots*)

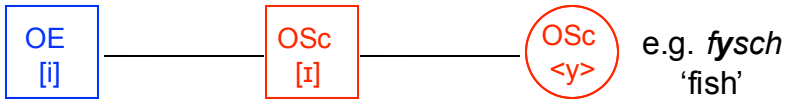
Benjamin Molineaux, Joanna Kopaczyk, Warren Maguire, Vasilis Karaiskos, Bettelou Los and Rhona Alcorn,

Funded by AHRC grant nº AH/L004542/1 for 2013-2018

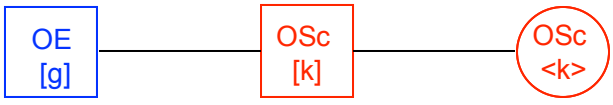
At the Angus McIntosh Centre for Historical Linguistics

- **Research Questions**
  - How can we account for the diversity of spelling attested in pre-16C Scots?
  - What can we learn about Scots phonology on the basis of those spellings?
- **Data**: c.1,250 'local documents' (c.400k words) written in Scots 1380-1500. From *A Linguistic Atlas of Older Scots* ('LAOS'; Williamson 2008)
- **Output**: A freely available, fully searchable, richly annotated corpus of triads

The FITS corpus of triads

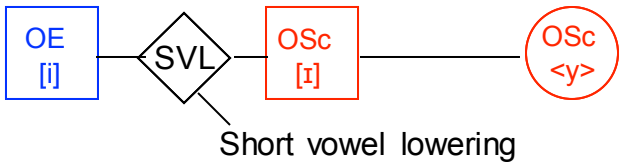


e.g. ***fysch*** 'fish'



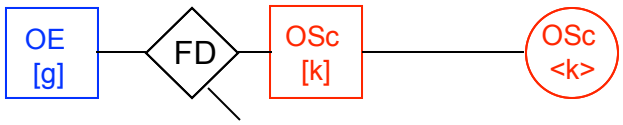
e.g. ***thynk*** 'thing'

The FITS corpus of changes



e.g. ***fy**sch*
'fish'

Short vowel lowering

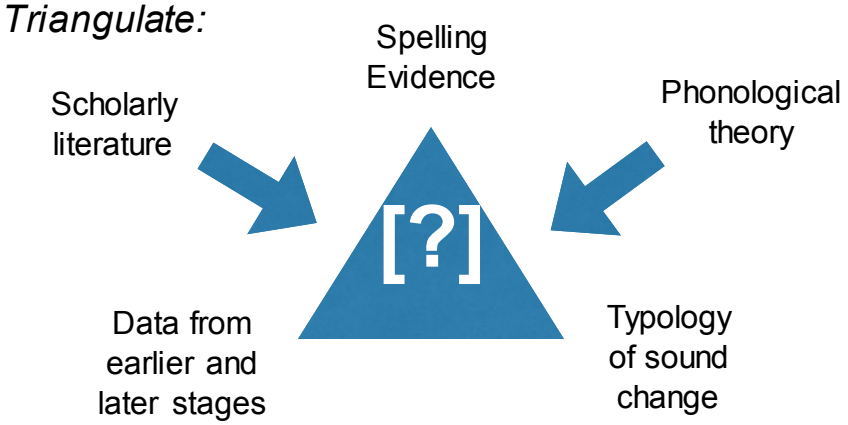


e.g. *thyn**k***
'thing'

Final devoicing

How do we reconstruct OSc sound values?

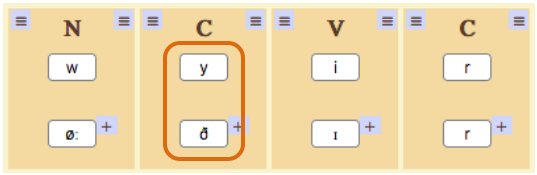
Triangulate:



Grapho-phonological parsing

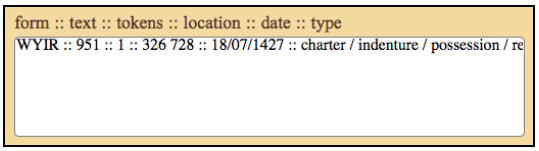
(Kopaczyk et al. 2018)

- Resolves word forms into units of spelling, e.g.
  - <wyir> = <w> | <y> | <i> | <r>
- Assign likely sound values



OSc spelling units

OSc sound units



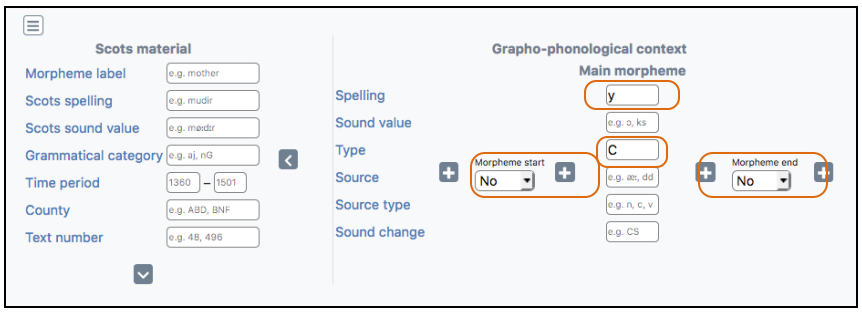
Linguistic context
word class, adjacent units

Extralinguistic info
date, location, text

Grapho-phonological parsing

(Kopaczyk et al. 2018)

- The database allows us to search for sounds and spellings by a number of linguistic and extra-linguistic parameters
  - all morpheme-internal consonantal uses of <y>



Grapho-phonological parsing

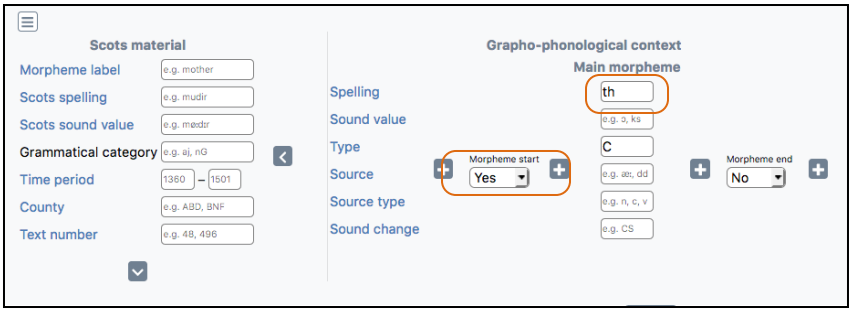
(Kopaczyk et al. 2018)

Form	Sound	Morpheme	Grammar	Texts Text ID–Year–County–Tokens	Total tokens
a±noyir	Root#-ø:ðɪr	other OED DSL	aj	146–1428–ELO–1	1
a±nuy(er)	Root#-ø:ð[][]	other OED DSL	aj	835–1400–FIF–1	1
brey(er)	bre:ð[][]	brother OED DSL	npl	83–1440–BWK–2	16
broy(er)	brø:ð[][]	brother OED DSL	n	61–1428–WLO–5	33
broy(er)sone	brø:ð[][]-±Root	brother OED DSL	n	955–1444–PTH–1	1
broy(er)sonis	brø:ð[][]-±Root	brother OED DSL	nG	955–1444–PTH–1	1
broy(er)son~	brø:ð[][]-±Root	brother OED DSL	n	332–1464–AGS–1	1
broyir	brø:ðɪr	brother OED DSL	n	465–1485–DMF–1	1
bruy(er)	brø:ð[][]	brother OED DSL	n	9521–1482–XAP–1	13
bruy(er)	brø:ð[][]	brother OED DSL	nG	827–1495–MLO–1	1
bruy(er)is	brø:ð[][]-#infl	brother OED DSL	nG	816–1493–MLO–1	1
bruyir	brø:ðɪr	brother OED DSL	n	81–1442–BWK–1	1
fay(er)	fað[][]	father OED DSL	n	325–1395–AGS–1	15
fayer	faðer	father OED DSL	n	744–1488–NRN–1	1

Grapho-phonological parsing

(Kopaczyk et al. 2018)

- all morpheme-internal consonantal uses of <th>



Grapho-phonological parsing

(Kopaczyk et al. 2018)

Form	Sound	Morpheme	Grammar	Texts Text ID–Year–County–Tokens	Total tokens
thotht	θɔxt	think OED DSL	vpp	9528–1487–XAP–3	4
thoucht	θɔxt	think OED DSL	vpt	66–1390–BWK–1	1
thr(e)	θr[]	three OED DSL	qc	1636–1444–ABD–1	2
thre	θreː	three OED DSL	qc	93–1384–FIF–1	229
thre()	θreː-#Root	three OED DSL	qc	1424–1446–AGS–1	10
thre()	θreː-#Root	three OED DSL	qo	1526–1457–PBL–1	1
three	θreː	three OED DSL	qc	768–1413–DNB–1	7
thyng	θɪŋg	thing OED DSL	npl	1513–1457–PBL–1	2
thyng	θɪŋg	thing OED DSL	n	86–1407–BWK–2	14
thyng"	θɪŋg-backStroke	thing OED DSL	n	325–1395–AGS–1	10
thyng(is)	θɪŋg-+abbrev/#infl	thing OED DSL	npl	125–1398–XDI–1	31
thynge	θɪŋg-e(abbrev)	thing OED DSL	n	123–1398–XDI–1	2
thyngeȝ	θɪŋg-e/#infl	thing OED DSL	npl	8001–1443–DURH–1	1

Sound-spelling mapping: Older Scots <þ, y, th>

"Northern system":

Jordan 1934, Benskin 1977, 1982, Stenroos 2004, Jensen 2012, Adamczyk 2016

Word-initially:

<þ/y> = /ð/ *they them there*

<th> = /θ/ *think through thousand*

Word-finally:

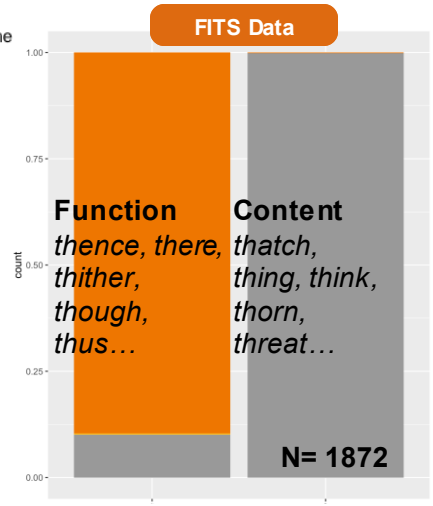
<th>= /θ/: *booth, north*

Word-medially:

<y> and <th> = /ð/ *other, brother*

This is claimed to be the result of gradual entry of <th> into the spelling system (Benskin 1977: fn 9).

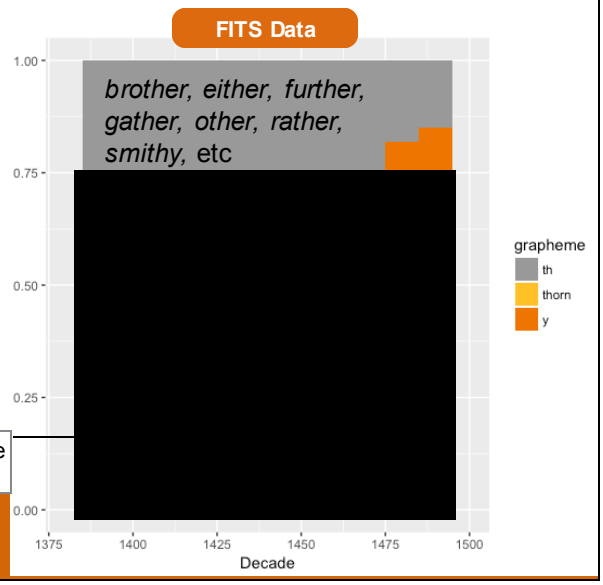
ba**th** > **th**ink > bro**th**er > **th**ere



AMC

Sound-spelling mapping: Older Scots <þ, y, th>

- Medial fricatives are etymologically voiced
- In early data we find far more <th> spellings
- By the end of the 15c, when our data is most abundant, <y> predominates
- <y> is growing as a marker of /ð/



density of attestations across the corpus period

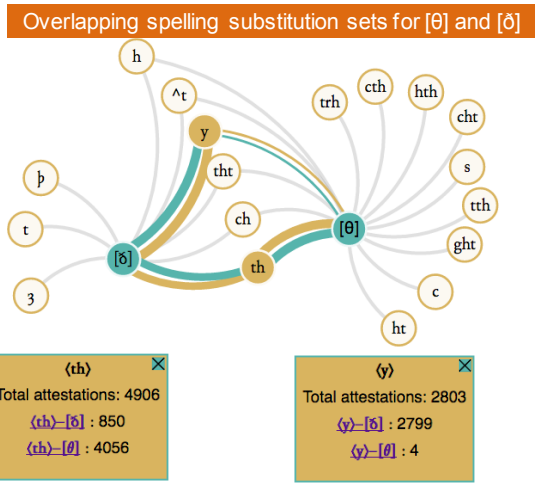
AMC

Sound-spelling mapping: Older Scots <þ, y, th>

- Older Scots displays an **emerging norm** that continues to develop throughout the period, at least for initial and medial position:
  - <y> is used for voiced contexts (initial function + medial)
  - <th> is used for voiceless contexts (initial content + final)
- The **<y>-for-voiced convention** appears to spread from initial to medial position, possibly a result of the initial spelling distinction between function and content words.

AMC

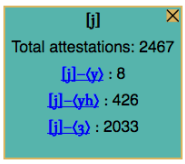
Spelling and sound substitution sets: Medusa

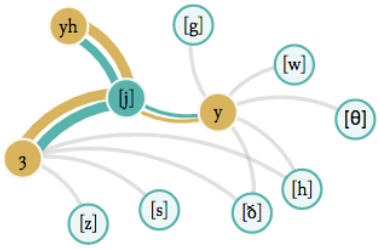


AMC

Spelling and sound substitution sets

- The use of <y> as a representation of the dental fricative has consequences for the system as a whole, with the **<yh> variant developing to represent /j/** in words like 'year' (alongside <y> and <ʒ>)



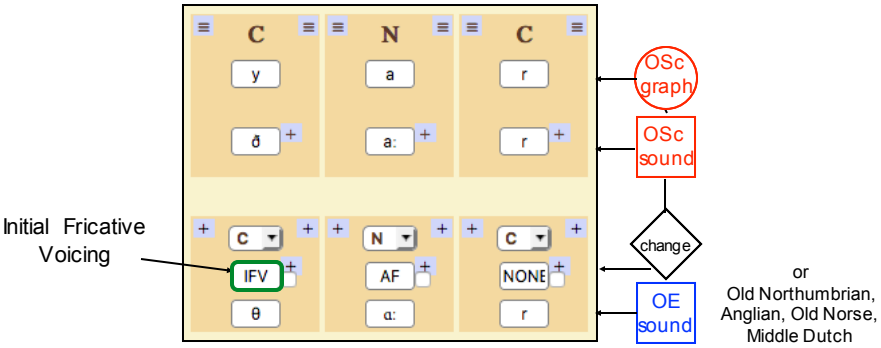


Overlapping sound substitution sets for <y>, <yh> and <ʒ>

AMC

Diachrony in the corpus

- In order to discover the changes that shaped the sound-system of Older Scots, we propose an etymological source for each item in the corpus.



- Each source segment (represented as a sound) is mapped on to a 15c Scots sound (where available) and a change mapping the one on to the other is proposed.

Diachrony in the corpus: IFV

- Sound changes are stored in a Corpus of Changes,
  - It gives a narrative for each change
  - Links to all proposed instances of the change

	IFV — Initial Fricative Voicing
Category	phonological
Description	The change operates in function words on the way from OE to ME and, we assume, MSc. See Lass (1992: 59) for ME, Johnston (1997: 98) for fricative phonemes in OSc. Luick (1914/1940: §703) dates it to the OE/ME transition, i.e. C11. *CoNE ((IFV))*

- The origins of spelling conventions are also accounted for in a corpus of spellings
  - It provides an account of the origins of a particular sound-spelling match in the history of the language

What will the FITS corpus do?

- **How to account for the diversity of spelling in pre-16C Scots?**
  - Grapho-phonological parsing: to link OSc spellings and sound
  - Etymological layer: to distinguish orthographic developments from phonological ones
- **What can we learn about Scots phonology from these spellings?**
  - Our corpus of triads identifies and quantifies:
    - relationships between OSc spellings and OSc sounds
    - relationships between OSc sounds & their etymological sources
    - the distribution of these relationships over time & space and within the linguistic system
  - Our Corpus of Sound Changes & Corpus of Spelling Changes narrate diachronic developments

Angus McIntosh Centre for Historical Linguistics

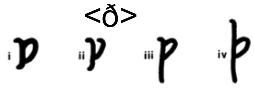
frlsu

Thanks!

More at www.amc.lel.ed.ac.uk/fits/

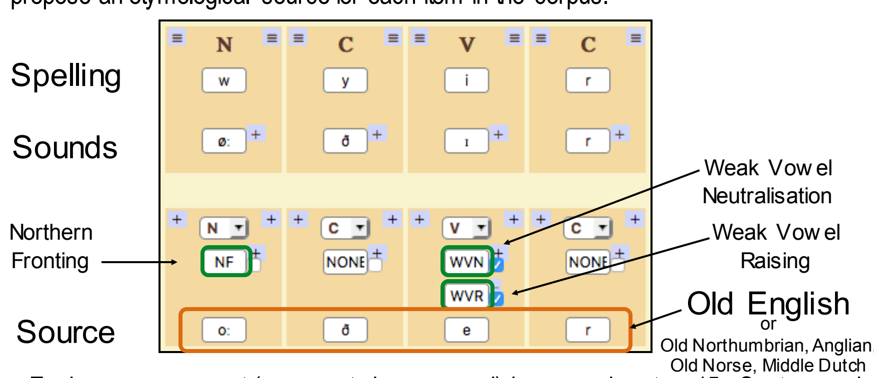
23 August, 2018

Sound-spelling mapping: Older Scots <þ, y, th>

	Spellings		Sounds	
Old English	<þ, ð> **interchangeable**	=	/θ/ → [ð] [θ]	between voiced segments: [oðer] elsewhere: [θiŋg, θere]
	<ð>  Benskin (1982)	=	/θ/ > /ð/ (phonemisation) /θ/	between voiced segments: /oðer/ word initially in function words: /ðere/ elsewhere: /θiŋg/
Older Scots	<þ> ~<y> + <th> **interchangeable?**	=	/θ/ /ð/	**th**ink, **th**igh, ba**th** **th**ere, **th**ey, bro**th**er

Diachrony in the corpus

- In order to discover the changes that shaped the sound-system of Older Scots, we propose an etymological source for each item in the corpus.



- Each source segment (represented as a sound) is mapped on to a 15c Scots sound (where available) and a change mapping the one on to the other is proposed.

Diachrony in the corpus: Northern Fronting

- Old English [o:] maps on to a number of spellings in OSc:

Total attestations: 4124

[ø:]–⟨eo⟩ : 2
[ø:]–⟨eu⟩ : 3
[ø:]–⟨i⟩ : 1
[ø:]–⟨o⟩ : 2111
[ø:]–⟨oi⟩ : 33
[ø:]–⟨oo⟩ : 4
[ø:]–⟨ou⟩ : 99
[ø:]–⟨oui⟩ : 1
[ø:]–⟨ov⟩ : 11
[ø:]–⟨ow⟩ : 92
[ø:]–⟨owy⟩ : 1
[ø:]–⟨oy⟩ : 32

[ø:]–⟨oyv⟩ : 2
[ø:]–⟨oyw⟩ : 1
[ø:]–⟨u⟩ : 1185
[ø:]–⟨ui⟩ : 9
[ø:]–⟨uv⟩ : 1
[ø:]–⟨uy⟩ : 30
[ø:]–⟨v⟩ : 411
[ø:]–⟨w⟩ : 83
[ø:]–⟨wi⟩ : 4
[ø:]–⟨y⟩ : 5
[ø:]–⟨yo⟩ : 1
[ø:]–⟨yv⟩ : 1
[ø:]–⟨yw⟩ : 1

- OSc <o> makes up about half of the attestations, but there are numerous other options
- We know from later reflexes and the literature, that [o:] normally fronted [ø:] (and raised [y:])
- There is no obvious spelling for [ø:] using the Roman alphabet

AMC

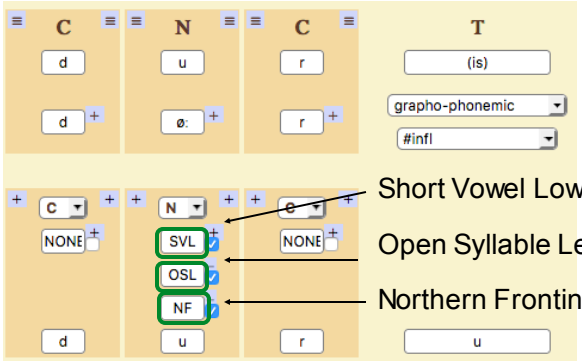
Diachrony in the corpus: Northern Fronting

- Following the literature, we postulate a change from [o:] > [ø:]: 'Northern Fronting'
- The change, doesn't only affect OE [o:], but also other OE elements that later join this category, such as stressed [u] in open syllables, such as in OE *duru* 'door'

15c sounds

CHANGES

OE sounds



AMC

Diachrony in the corpus: Northern Fronting

- Sound changes are stored in a Corpus of Changes,
  - It gives a narrative for each change
  - Links to all proposed instances of the change

	HL — Homorganic Lengthening
Category	phonological
Description	Homorganic Lenthening (HL) refers to the lengthening of short vowels before clusters of two consonants with the same place of articulation, the first of which is a sonorant, the second of which is typically a voiced obstruent or nasal. This change is traditionally dated to the second half of the [illegible]th and first half of the 9th centuries (Jordan 1934: 39, Minkova 2014: 165; see also Jespersen (DATE: 118-120), Lass (CHEL2, 1999: 71-72) and Luick (DATE: 474-477), and Aitken & MacAfee (2002: 6-9) and Johnston (1997a: 65-66) specifically for accounts of the change in the history of Scots). The clusters ostensibly involved in HL were: *ld* (e.g. OE *cild* 'child'), *mb* (e.g. OE *climban* 'climb'), *nd* (e.g. OE *blind* 'blind'), *ng* (e.g. OE *lang* 'long'), *rd* (e.g. OE *beard* 'beard'), *rl* (e.g. OE *eorl* 'earl'), *rn* (e.g. OE *bearn* 'bairn, child'), *rð* (e.g. OE *eorðe* 'earth'). When a third consonant followed, HL was blocked (e.g. OE *cildru* 'children').

THIS WILL BE THE ENTRY FOR NORTHERN FRONTING

- The origins of spelling conventions are also accounted for in a corpus of spellings
  - It provides an account of the origins of a particular sound-spelling match in the history of the language