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

OE [i] OSc [i] OSc &lt;y&gt; e.g. *fish*

OE [g] OSc [k] OSc &lt;k&gt; e.g. *think*
The FITS corpus of changes

Short vowel lowering

Final devoicing

How do we reconstruct OSc sound values?

**Triangulate:**

- Scholarly literature
- Spelling Evidence
- Phonological theory
- Data from earlier and later stages
- Typology of sound change
Grapho-phonological parsing (Kopaczyk et al. 2018)

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

**Linguistic context**
- Word class, adjacent units

**Extralinguistic info**
- Date, location, text

**Grapho-phonological context**
- 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)

- all morpheme-internal consonantal uses of <th>
Grapho-phonological parsing (Kopaczky et al. 2018)

<table>
<thead>
<tr>
<th>Form</th>
<th>Sound</th>
<th>Morpheme</th>
<th>Grammar</th>
<th>Texts</th>
<th>Total tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>thet</td>
<td>8oth</td>
<td>think OED DSL</td>
<td>vpp</td>
<td>9528-1487-XAP-3</td>
<td>4</td>
</tr>
<tr>
<td>thought</td>
<td>8ought</td>
<td>think OED DSL</td>
<td>vgt</td>
<td>66-1369-BWK-1</td>
<td>1</td>
</tr>
<tr>
<td>th(e)</td>
<td>8(e)</td>
<td>three OED DSL</td>
<td>gc</td>
<td>1636-1444-ABO-1</td>
<td>2</td>
</tr>
<tr>
<td>th(e)</td>
<td>8e</td>
<td>three OED DSL</td>
<td>qf</td>
<td>90-1364-FIF-1</td>
<td>229</td>
</tr>
<tr>
<td>three</td>
<td>8ree</td>
<td>three OED DSL</td>
<td>qf</td>
<td>1456-1446-AGS-1</td>
<td>10</td>
</tr>
<tr>
<td>three</td>
<td>8ree-8Root</td>
<td>three OED DSL</td>
<td>qf</td>
<td>1536-1457-PBL-1</td>
<td>1</td>
</tr>
<tr>
<td>three</td>
<td>8ree</td>
<td>three OED DSL</td>
<td>qf</td>
<td>766-1413-DNB-1</td>
<td>7</td>
</tr>
<tr>
<td>thng</td>
<td>8ng</td>
<td>thing OED DSL</td>
<td>npl</td>
<td>1513-1457-PBL-1</td>
<td>2</td>
</tr>
<tr>
<td>thng</td>
<td>8ng-backStroke</td>
<td>thing OED DSL</td>
<td>n</td>
<td>95-1467-BWK-2</td>
<td>14</td>
</tr>
<tr>
<td>thng</td>
<td>8ng-8abbrev</td>
<td>thing OED DSL</td>
<td>npl</td>
<td>325-1395-AGS-1</td>
<td>10</td>
</tr>
<tr>
<td>thng</td>
<td>8ng-8abbrev</td>
<td>thing OED DSL</td>
<td>npl</td>
<td>125-1398-XDI-1</td>
<td>31</td>
</tr>
<tr>
<td>thng</td>
<td>8ng-a8abbrev</td>
<td>thing OED DSL</td>
<td>n</td>
<td>125-1398-XDI-1</td>
<td>2</td>
</tr>
<tr>
<td>thng</td>
<td>8ng-a8abbrev</td>
<td>thing OED DSL</td>
<td>npl</td>
<td>8001-1443-DURH-1</td>
<td>1</td>
</tr>
</tbody>
</table>

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

“Northern system”:

Word-initially:
<p/y> = /ð/  <th> = /θ/:
*they* *think*
*them* *through*
*there* *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).

- *bath* > *think* > *brother* > *there*
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/.

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.
Spelling and sound substitution sets: Medusa

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 /lj/ in words like 'year' (alongside <y> and <ʒ>)
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 onto 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

- 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

AMC

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

23 August, 2018
Sound-spelling mapping: Older Scots <þ, y, th>

<table>
<thead>
<tr>
<th>Spellings</th>
<th>Sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old English: &lt;þ, ð&gt;</td>
<td>/ð/ → [ð]</td>
</tr>
<tr>
<td>=</td>
<td>[ð]</td>
</tr>
<tr>
<td>Old Scots: &lt;þ&gt; ~&lt;y&gt;</td>
<td>/θ/</td>
</tr>
<tr>
<td>+ &lt;th&gt;</td>
<td>/ð/</td>
</tr>
</tbody>
</table>

*Interchangeable?

Benskin (1982)

- Between voiced segments: [ðer]
- Elsewhere: [þing, ðere]

- Certain voiced segments: /ðer|
- Word initially in function words: /ðere|
- Elsewhere: /ðing|

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 onto a 15c Scots sound (where available) and a change mapping the one onto the other is proposed.
Diachrony in the corpus: Northern Fronting

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

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

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

<table>
<thead>
<tr>
<th>Category</th>
<th>phonological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Homorganic Lengthening (HL) refers to the lengthening of most vowels before clusters of two consonants with the same place of articulation. This Fronting process is often dated to 1000 AD. A notable example is from the historical account by Malcolm (2015), 474-477, and Kilian &amp; MacFhionnlaoch (2002). See also (Burnett 1996: 86-88) specifically for accounts of the change in the history of Scots. The clusters historically involved in HL were: (e.g. OE old 'olde'), (e.g. OE old 'old'), (e.g. OE old 'old'), (e.g. OE old 'old'), (e.g. OE full 'ful'), (e.g. OE full 'full'), (e.g. OE full 'full'), (e.g. OE full 'full'), (e.g. OE full 'full'). When a third consonant followed, HL was blocked (e.g. OE old 'children').</td>
</tr>
</tbody>
</table>

- 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

AMC