Social embedding of neologisms in early English correspondence

Tanja Säily, Eetu Mäkelä & Jukka Suomela
STRATAS project

• *Interfacing structured and unstructured data in sociolinguistic research on language change* (Academy of Finland, DIGIHUM, 2016–2019)

• Our subproject: *Social embedding of neologisms in early English correspondence*
  – Previous research: mostly lexicographical data (OED), bias towards well-known authors
  – *Corpora of Early English Correspondence* (CEEC): wide social spectrum, speech-like genre
CEEC

- **Personal letters**, c. 1400–1800
  - 1,180 writers, 11,713 letters, 5.2 million words
  - Compiled for historical sociolinguistics: metadata on letters, writers, recipients (e.g. gender, social rank)
- Compiled by T. Nevalainen, H. Raumolin-Brunberg et al. at the University of Helsinki
  - Based on published editions of letters
- Published versions:
  - *Sampler*, 1998 (0.5 Mw, 1418–1680)
  - *Parsed*, 2006 (2 Mw, 1410–1681)
Research questions

1. **Who** are the innovators? Which social groups do they represent?
2. **How do the new words spread** socially, geographically and diachronically?
3. **Which semantic domains** do the neologisms represent?
4. **Why** are the neologisms created and established? Can they be linked to:
   - Specific historical events?
   - Changes in culture & society?
   - Social meanings?
Challenges

• **Identification** of neologisms
  – Spelling variation…

• Automatic retrieval of related **lexicographical** data
  – *Oxford English Dictionary*: e.g. first attestation dates, etymologies
  – *Historical Thesaurus*: time-sensitive semantic domains

• **Interface** for pruning the possible neologisms found
Getting started: case studies

• Based on previous work: variation in the productivity of nominal suffixes -ity, -er

• **Material:** *Corpus of Early English Correspondence Extension*
  – Late C18 section, 1760–1800

• **FiCa** interface for classifying data, retrieving OED first attestation dates
  – Developed by Eetu Mäkelä
<table>
<thead>
<tr>
<th>FiCa</th>
<th>RLR (3)</th>
<th>RLR (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>roller (3)</td>
<td>rollers</td>
<td>yes</td>
</tr>
<tr>
<td>roller</td>
<td>roller</td>
<td>yes</td>
</tr>
<tr>
<td>roller</td>
<td>roller</td>
<td>yes</td>
</tr>
<tr>
<td>ruler</td>
<td>ruler</td>
<td>unclear</td>
</tr>
<tr>
<td>rem-ber</td>
<td>rem-ber</td>
<td>no</td>
</tr>
<tr>
<td>rambler (2)</td>
<td>rambler</td>
<td>unclear</td>
</tr>
<tr>
<td>rambler</td>
<td>rambler</td>
<td>no</td>
</tr>
<tr>
<td>rem[j]ber</td>
<td>rem[j]ber</td>
<td>no</td>
</tr>
<tr>
<td>remember (2)</td>
<td>remember</td>
<td>no</td>
</tr>
<tr>
<td>remember</td>
<td>remember</td>
<td>no</td>
</tr>
<tr>
<td>remember (2)</td>
<td>remember</td>
<td>no</td>
</tr>
<tr>
<td>members</td>
<td>members</td>
<td>no</td>
</tr>
<tr>
<td>remembrer</td>
<td>remembrer</td>
<td>no</td>
</tr>
<tr>
<td>remembrancer</td>
<td>remembrancer</td>
<td>yes</td>
</tr>
<tr>
<td>remainder (9)</td>
<td>remainder</td>
<td>no</td>
</tr>
</tbody>
</table>

This is a late invention; but as it requires more workmanship than is easy to be had here, I think cast iron rollers of small diameter which the seed is to be subjected to before it is exposed to the pressure of the great stone...
Case 1: -ity

Corpora of Early English Correspondence, 1760–1800
Introduction

• -ness & -ity: derive nouns from adjectives
  – e.g. generous → generosity / generosity
  – -ness native, -ity borrowed from French & Latin
    → connotations associated with those languages
    (e.g. polite society, learned, scientific)

• Previous research (Säily 2014): productivity of
  -ity increases throughout C17–18, perhaps led by
  male professionals

• End of C18: -ity highly productive –
  what kinds of neologisms by whom?
15 neologisms found, 1760–1800

• Working definition of neologism: corpus attestation max 100 years after OED first attestation date
  – 3 antedatings, 3 not in OED at all

• authenticity, cleverality (1778<1828), comicality, conviviality (1783<1791), coxcombicality, foxity (1788), Germanity (1788<1821), impracticability, intrepidity, irritability, oddity, respectability, ridiculability (1776), scoundrellity (1761), versatility
Who are the innovators?

• **Thomas Twining**, clergyman, scholar (c.1734–1804)
  – coxcombicality, **foxity**, Germanity 1788, ridiculability 1776

• **Charles Burney**, musician, author (1726–1814)
  – **conviviality** 1783, versatility 1782
  – Daughter **Fanny Burney** (author, 1752–1840) uses **conviviality** 1799

• **David Garrick**, actor, playwright (1717–1779)
  – comicality 1767, **scoundrelly** 1761

• **Ignatius Sancho**, author (1729?–1780)
  – cleverality 1778
How do they use the neologisms?

... there is not among any set of people such a comfortable scratch-back confederacy as among those old ruin-diggers [antiquarians & old-English grubbers]. Is it the consciousness of enemies & scoffers all round them – the sour’d feeling of *ridiculability* – that draws the knot closer and forms the phalanx, back to back, for mutual scrubbing & defence?

Thomas Twining to Charles Burney, 1776

It is a most infamous design, & I desir’d Churchill would Let Thornton know of it, which he will do immediatly, & prevent their *Scoundrillity* by some humourous Paragraph ...

David Garrick to George Colman, 1761
How do neologisms spread?

I have always thought, that in many particulars his equal was not to be found – his wit, learning, taste, penetration; & when well, his conviviality, pleasantry, – & kindness of heart to me & mine, will ever be thought of, with the most profound & desponding regret!

Charles Burney to Fanny Burney, 1783?

The account of the Play meriting little attention indeed – I am much pleased at your independent establishment of conviviality at Burlington House.

Fanny Burney to Charles Burney, 1799
-ity: answers to RQs

1. **Innovators**: creative men in their 40s–50s, professionals / other non-gentry
2. Neologisms **spread** in a social network of peers writing in a similar style
3. They often describe **human attributes**
4. They also carry **social meaning**: designed to amuse / impress recipient, emphasise in-group membership
Case 2: -er

Corpora of Early English Correspondence, 1760–1800
Nominal suffixes -er and -or

• Typically derive **agentive** or instrumental nouns from verbs: *driver, governor, filler*
  – Also: person ‘concerned with N’ / ‘living in N’
• -or: Latinate variant of -er
  – Pronounced identically (*adviser/advisor*)
    → treated as a single suffix here
    (Plag 2003: 89; Bauer 2001: 199–203)
Previous research (CEEC, C18)

• **Productivity** of -er increases over time
  – Stylistic change or continued semantic expansion? (Säily 2014, -ity; Dalton-Puffer 1994)

• Men writing to their close friends overuse -er
  – Less stable relationship a trigger for productivity? (cf. Wolfson 1990; Säily 2014, -ity)

• End of C18: -er highly productive – what kinds of neologisms by whom?
25 neologisms found, 1760–1800

• **Describing people:** *absconder*, *blubberer* (1782<1786), *commemorator* (1784<1856), *completer*, *complimenter*, *dangler*, *outsider* (1800=OED), *schemer*, *seceder*, *spiter* (1790<1847), *swindler*
  – **Occupations:** *(shirt-)*airer, gambler, hairdresser, *(China-)*piecer, smuggler

• **Connected to places:** *Chiswicker*, *Madrasser*, *Norfolker*, *Turnham-Greener* (none in OED)

• **Things:** *cutter* ‘boat’, *ventilator*

• **Other:** *brightener*, *plumper* ‘lie’ (1776=OED), *winterer* (1784<1795)
Who are the innovators?

• Thomas Twining, clergyman, classical scholar (c.1734–1804)

• Hester Piozzi, writer (1741–1821)
  – *seceder, spiter* 1790

• Jane Austen, novelist (1775–1817)
  – *outsider* 1800

• George Culley, agriculturist (c.1735–1813)
  – *Norfolker, winterer* 1784
How do they use the neologisms?

As to myself, *Cecilia* has done just what she pleas’d with me: I laughed, & cried (for I am one of the **blubberers**) when she bade me.

Thomas Twining to Charles Burney, 1782

... it makes me laugh when I think how the **Spiters** told us that *Siddons had lost all her Popularity* ...

Hester Piozzi to Charlotte Lewis, 1790
How do neologisms spread?

• All of the -er antedatings are *hapax legomena* in the corpus → no hard evidence
• However, all of them (like -ity) occur in letters written to **close friends or family**
  — Spread in the social networks of innovators?
  — Not all may have spread (depending on status of innovator in the network?), could have been recoined at a later date
-er: answers to RQs

1. Innovators: men *and* women, mostly in their 40s–50s, professional / other non-gentry
2. Neologisms *spread* in a social network of peers writing in a similar style
3. They often describe *people*
4. They also carry *social meaning*: designed to amuse recipient, emphasise in-group membership
   – Some reflect *innovations in society/(agri)culture*
Conclusion
Case studies

• Late C18 neologisms mostly produced by the emerging **middle class**, reflect **involved style** of letter-writing (affective, interpersonal)
  – **ridiculability, spiter**

• Most innovative social groups **vary by affix**
  – **-ity**: men in creative professions, **-er**: both men and women, professionals and other non-gentry

• **Purpose** of innovation may vary by affix
  – **-er** also used for societal innovations
DynLex in a sociohistorical corpus

• Following the spread of **individual words across individual people** may prove impossible
  – CEEC doesn’t include complete correspondences
  – Sparse data: 5 million words over 400 years

• What to do instead?
  – Supplement with other datasets?
  – Track neologisms at the level of **semantic domains** (Alexander et al. 2015) and **social groups**?
    • Do neologisms in a particular semantic domain originate with a specific social group, spread to others?
Beyond one-off case studies

• Semi-automatically filter CEEC corpus words to neologisms
• Explore word spread and social factors interactively
Discovering -ER in CEEC

- er(e), -ar(e), -or(e), -our(e), -owr(e), -ur(e), -r + plural, possessive…

- \S*(((\[rR\]|\[eEoO]\~)(=\?|=*\[eE]=\?|[^=\~]*)[^EiIyY]\[^\~]*[^sSzZ][^\~]*)\[^\~]*[^a-zA-Z\~+=])

- 6800 candidate types, 300 000 appearences
- 5080 types out of 6800 irrelevant after manual study
- 153 types out of 6800 needed further study
  - 11768 individual uses
Semi-automatically filtering CEEC corpus words to neologisms

• Filter candidates by first attestations in OED, EEBO(-TCP), ECCO, BURNNEY, BLN, TDA
• Problems: spelling variation, OCR errors in comparison corpora
• Solutions:
  – match based on multiple algorithms (phonetic, lemma, …)
  – calculate match confidence (e.g. factor in total word frequency in all corpora)
  – use grouping and sorting functionalities of FiCa to efficiently verify candidates and spot errors
Semi-automatically filtering CEEC corpus words to neologisms

- If still too many candidates, create subcorpora of more manageable sizes
  - by dates
  - by social classes of interest
  - by later behaviour of the words in the combined corpora (e.g. only those that do appear later?)
- CEEC types in BLN+BURNEY+ECCO
  - ~58 000 types have less than 1000 appearances
  - ~36 000 types have less than 100 appearances
  - ~17 000 less than 10, ~2 500 appear once, ~6 700 don’t appear at all
References


• CEEC = *Corpora of Early English Correspondence*. Compiled by T. Nevalainen, H. Raumolin-Brunberg et al. at the University of Helsinki. [http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/](http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/)


