NEW METHODS FOR ANALYSING SUFFIX COMPETITION ACROSS REGISTERS

HOW -ITY GAINED GROUND ON -NESS IN EARLY MODERN ENGLISH

Tanja Säily, Paula Rodríguez-Puente and Jukka Suomela
EARLY MODERN ENGLISH

• Large-scale expansion of vocabulary
  • Borrowing + highly productive use of word-formation processes
  • Vernacularization, standardization (Nevalainen 1999: 358; Nevalainen & Tieken-Boon van Ostade 2006: 301–303)

• Middle English: intensive borrowing, decline of native affixal system
  • Beginning of EModE period: still a lot of freedom of choice in affix usage, parallel derivatives
RIVAL SUFFIXES: -ITY AND -NESS

• **Nominal suffixes**, usually derive abstract nouns from adjectives
  • E.g. *productive* → *productivity* or *productiveness*

• -ness native, -ity borrowed from French (+ Latin) in Middle English
  • More sociolinguistic variation in the productivity of -ity (Säily 2014); prestige, learnedness

• Early Modern English: *register variation* in the use of the suffixes?
RODRÍGUEZ-PUENTE (2020)

• Analysis of -ness and -ity in 17 EModE registers
  • More comprehensive than previous studies (e.g. Cowie 1998)

• Material from 3 corpora
  • Corpus of English Dialogues (CED, 1560–1760)
  • Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME, 1500–1710)
  • EModE section of the Corpus of Historical English Law Reports, (CHELAR, 1535–1999)

• Hypothesis: -ity more learned → more common in specialized written registers
### REGISTERS STUDIED

(RODRÍGUEZ-PUENTE 2020; CF. CULPEPER & KYTÖ 2010)

<table>
<thead>
<tr>
<th>Speech-related</th>
<th>Informal</th>
<th>Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech-like</strong></td>
<td>Diaries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Letters, private</td>
<td></td>
</tr>
<tr>
<td><strong>Speech-based</strong></td>
<td>Trial proceedings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Witness depositions</td>
<td></td>
</tr>
<tr>
<td><strong>Speech-purposed</strong></td>
<td>Drama</td>
<td>Sermons</td>
</tr>
<tr>
<td><strong>Writing-based and writing-purposed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Auto)biography</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Travelogue</td>
<td></td>
</tr>
</tbody>
</table>

Suffix competition across registers / Säily, Rodríguez-Puente, Suomela 2021-06-03
RESULTS
(RODRÍGUEZ-PUENTE 2020)

• *-ity gains ground on -*ness in all registers!
  • Change seems to start from formal written registers, spread towards speech-related ones

• Related to a general trend towards a more literate style in the 18th century?
  • Biber & Finegan (1997), McIntosh (1998: 23–24)

• More statistical evidence needed → collaboration
  • Rodríguez-Puente, Säily & Suomela (forthcoming)
MORPHOLOGICAL PRODUCTIVITY

• The readiness with which an element enters into new combinations (Bolinger 1948)
• Quantitative measures (e.g. Baayen 1993; Cowie & Dalton-Puffer 2002):
  • Number of different words containing the morpheme in a corpus (types)
  • Number of types occurring only once in the corpus (hapax legomena)
  • Number of types not occurring in previous periods (new types)
• Problem: Difficult to compare across (sub)corpora
  • Different amounts of data from different periods & registers
  • Type-based measures grow nonlinearly with corpus size → normalization not justifiable

![Graph of running words vs. types](image-url)
Running words (millions) vs. -ity

- Women
- Men
• Compare each subcorpus with subcorpora of equal size, randomly sampled from the corpus as a whole

• Automatically provides a measure of statistical significance

• Problems:
  • Comparisons over time still difficult; x-axis = corpus size, not time period
  • Only measures variation within a morpheme, not between morphemes

FURTHER ISSUE: PERIODIZATION

- Combining data from corpora covering partly different periods of time (+ registers)
  
  - Corpus of English Dialogues (CED), 1560–1747
  - Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME), 1500–1719
  - Corpus of Historical English Law Reports (CHELAR), 1544–1748
TOWARDS A SOLUTION

• Force **time on the x-axis** and see what it requires from the method

• Use a **sliding window for time periods** to combine the corpora

• Compare competing morphemes as if they formed a **linguistic variable**
  • Calculate proportion of *-ity* types out of all *-ity* and *-ness* types
• Corpora: CED + PPCEME
• Blue = writing-based & purposed, orange = speech-related registers
• 100-year sliding window, 25-year intervals

• Problems:
  • Different amounts of data from registers \(\rightarrow\) comparability?
  • Statistical significance?
FOCUS ON A SPECIFIC PERIOD…

- 1600–1699
FOCUS ON A SPECIFIC PERIOD…

- 1600–1699
- Return corpus size to x-axis
- Compare each register with random subcorpora of the same size sampled from the entire corpus
  - Proportion of -ity statistically significantly high in writing-based & purposed registers but still low in speech-related registers
DO THE SAME FOR ALL PERIODS IN BOTH REGISTERS…
RESULTS

- Proportion of -ity increases over time
- **Writing-based and purposed registers**: rapid increase from late 16th century onwards
- **Speech-related registers**: lagging behind until late 17th century
- 1650–1749: proportion of -ity close to 50%
- Potentially related to a general trend towards a more literate style in the 18th century (Biber & Finegan 1997; McIntosh 1998)
EXTREME EXAMPLE: LAW

- Corpora: CED + PPCEME + CHELAR
- Proportion of -ity statistically significantly high in “Law” register (statutes from PPCEME, law reports from CHELAR) throughout the Early Modern English period
- Did the change start here?
- Might reflect an overall preference for Romance vocabulary in the register since Middle English
CONCLUSIONS

- **Register a key factor** in the competition of *-ness* and *-ity* in the Early Modern period
  - Rise of *-ity* starts in formal written registers, spreads towards speech-related ones
- New methods → more **reliable & accurate results**
  - Enable investigation of register, social & intralinguistic factors in competing processes
  - Method of periodization usable in any diachronic study
- Robust statistical methods tend to require a lot of data → **larger corpus needed** to:
  - Analyse each individual register
  - Consider productivity based on hapax legomena or new types
REFERENCES


