



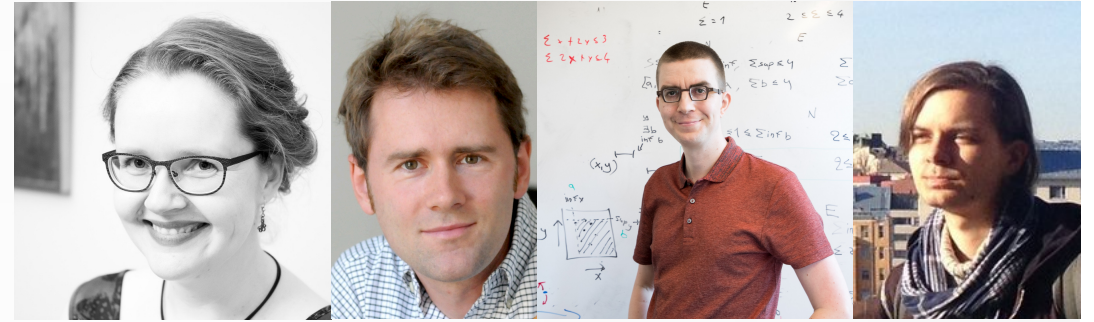
# NEW APPROACHES TO INVESTIGATING CHANGE IN DERIVATIONAL PRODUCTIVITY

Tanja Säily · Martin Hilpert · Jukka Suomela  
Helsinki · Neuchâtel · Aalto University





# HISCOP PROJECT

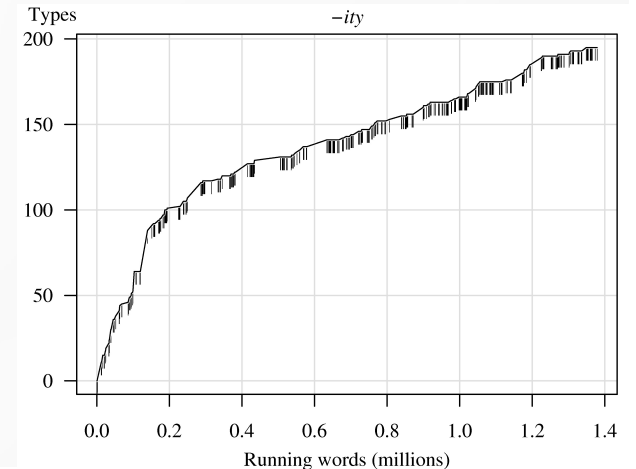


- *Historical Sociolinguistics Meets Construction Grammar: The Case of Productivity in English*
  - Academy of Finland, 2020–2023
  - Funded researcher: **Tanja Säily**
  - Collaborators: **Martin Hilpert**, **Jukka Suomela**, Florent Perek, Turo Vartiainen
  - Student assistant (2020): **Lassi Saario**
- Aim: extend CxG by drawing on historical sociolinguistics
  - What do speakers have to know to be able to use a language?  
Social aspects largely missing so far
  - Focus on productivity of constructions in historical text corpora



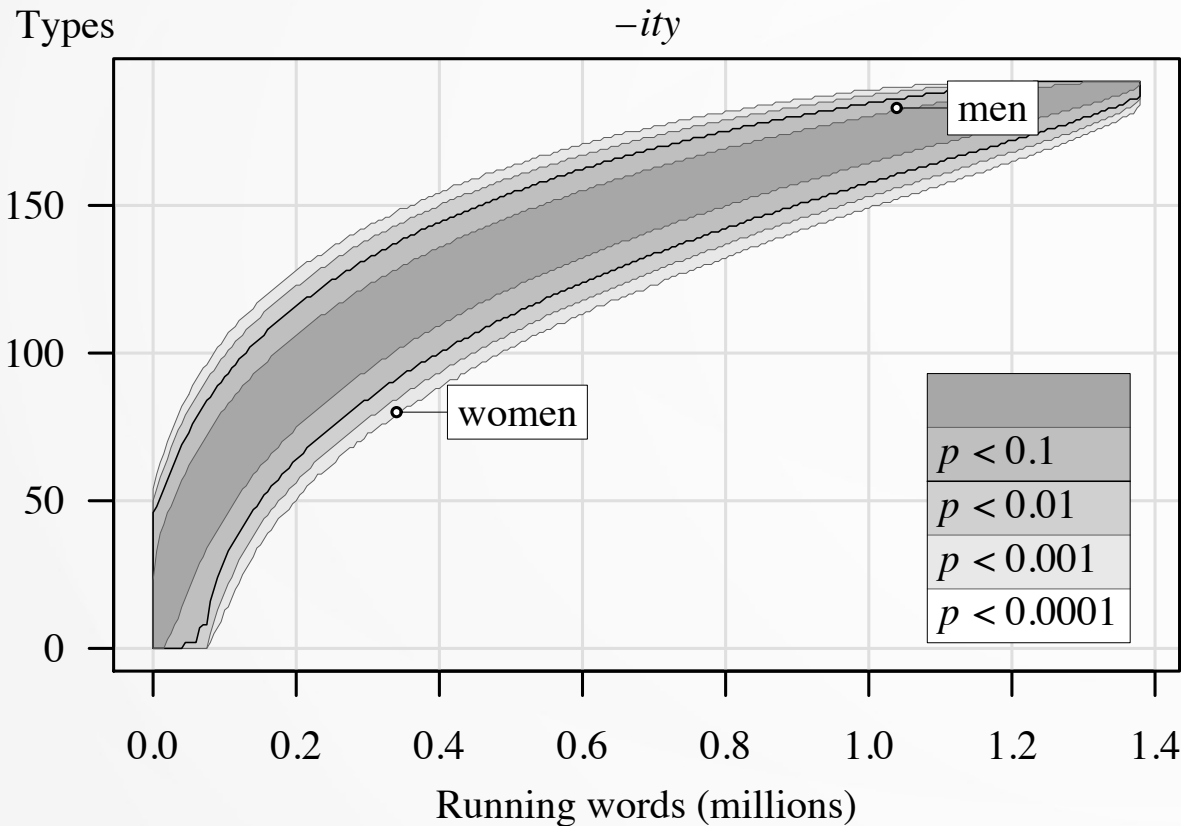
# MORPHOLOGICAL PRODUCTIVITY

- 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 & groups
  - Type-based measures grow nonlinearly with corpus size  
→ **normalization not justifiable**





# SÄILY & SUOMELA (2009, 2017): PARTIAL SOLUTION



- 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

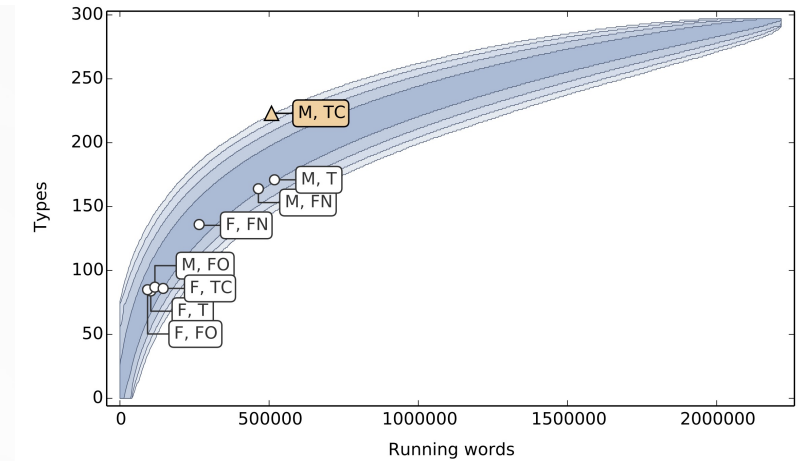


# ***-ITY AND -NESS***

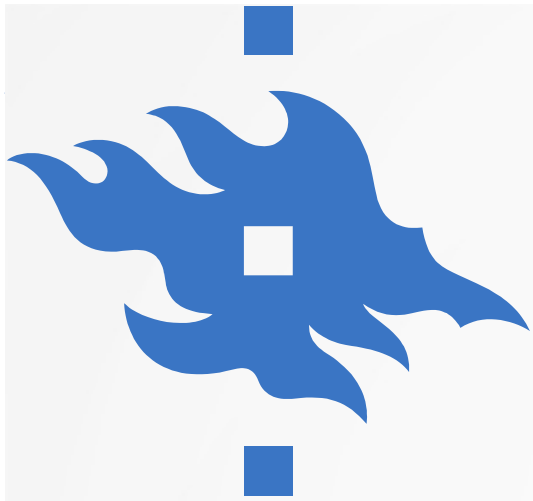
- **Nominal suffixes**, usually derive abstract nouns from adjectives
  - e.g. *productive* → *productivity* or *productiveness*
- *-ness* native, *-ity* borrowed from French (+ Latin) in Middle English
  - More sociolinguistic variation in the productivity of *-ity* (Säily 2014); prestige, learnedness
- Early Modern English: large-scale expansion of vocabulary
  - *-ity* gains ground on *-ness* in all registers, starting from written registers and spreading towards speech-related ones
    - Rodríguez-Puente (2020); Rodríguez-Puente et al. (submitted)



# ***-ITY* AND *-NESS* IN C17–18 PERSONAL LETTERS**



- Säily (2014): **external** factors
  - Productivity of *-ity* increases, *-ness* remains stable (*Corpora of Early English Correspondence*, type frequencies)
  - Gender: women lag behind in the use of *-ity* in C17, difference disappears in C18
    - Exception: difference remains in letters to close friends (cf. Wolfson 1990)
- Now: analyse suffix competition (cf. Rodríguez-Puente et al. submitted), add **internal** factors
  - Hilpert (2013): a number of language-internal factors connected to change in the productivity of the *V-ment* construction (*Oxford English Dictionary*, 1250–2000)
    - We will analyse some of the same factors



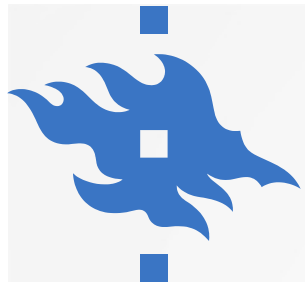
# SUFFIX COMPETITION



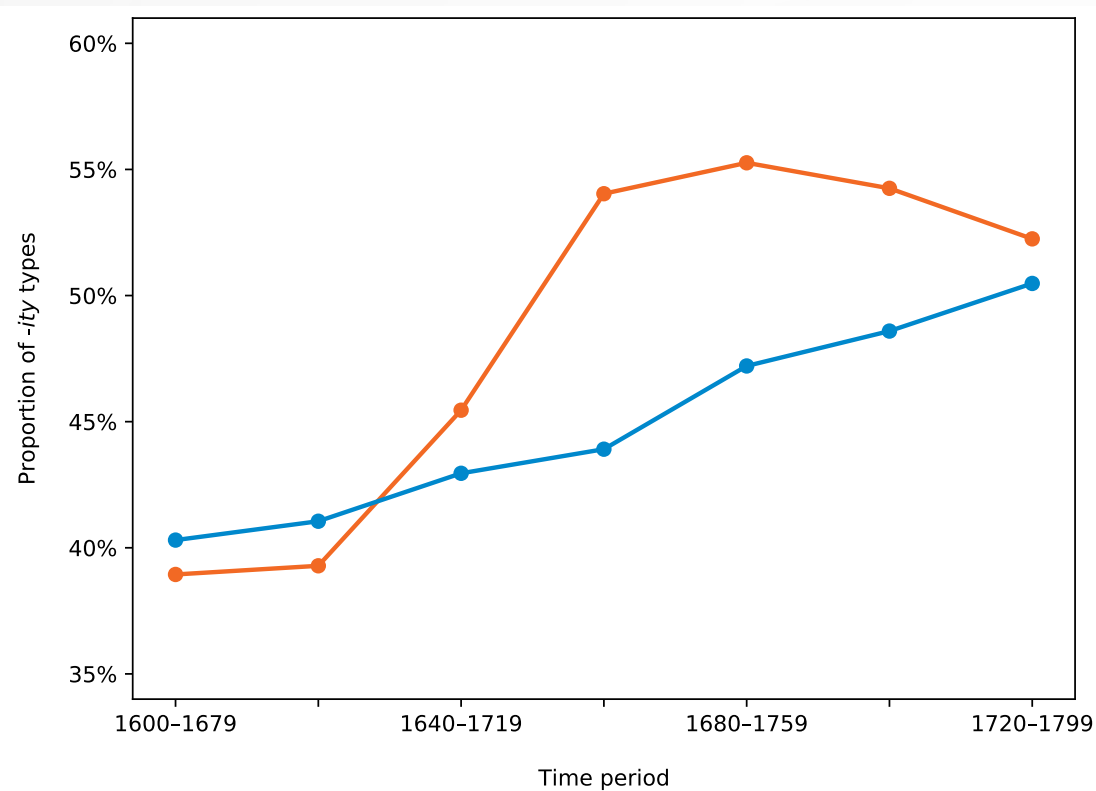
# ANALYSING SUFFIX COMPETITION

- **Problems with existing method:**
  - Comparisons over time difficult; x-axis = corpus size, not time period
  - Only measures variation within a morpheme, not between morphemes
- Towards a solution:
  - Force **time on the x-axis** and see what it requires from the method
  - Compare competing morphemes as if they formed a **linguistic variable**
    - Calculate proportion of *-ity* types out of all *-ity* and *-ness* types





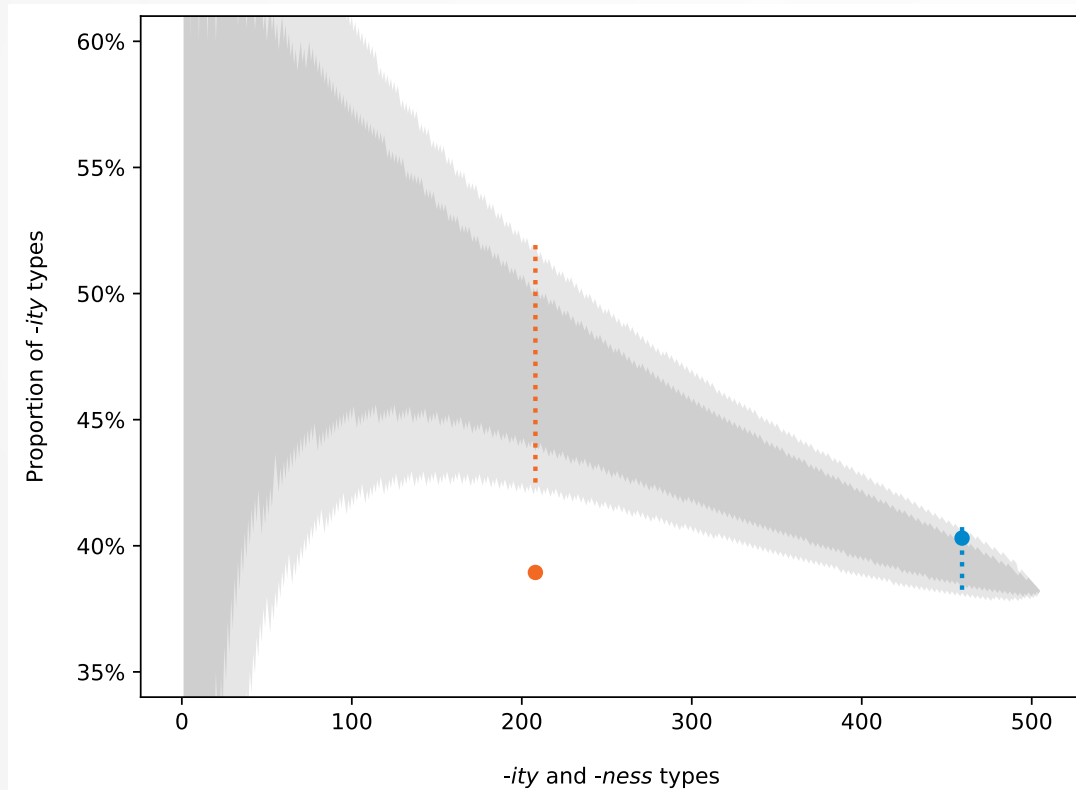
# FIRST ATTEMPT



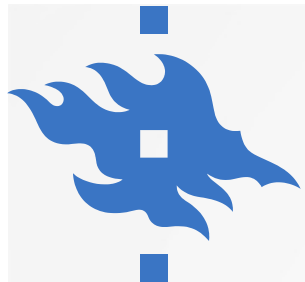
- **Blue** = men, **orange** = women
- 80-year sliding window, 20-year intervals
- **Problems:**
  - Different amounts of data from genders → comparability?
  - Statistical significance?



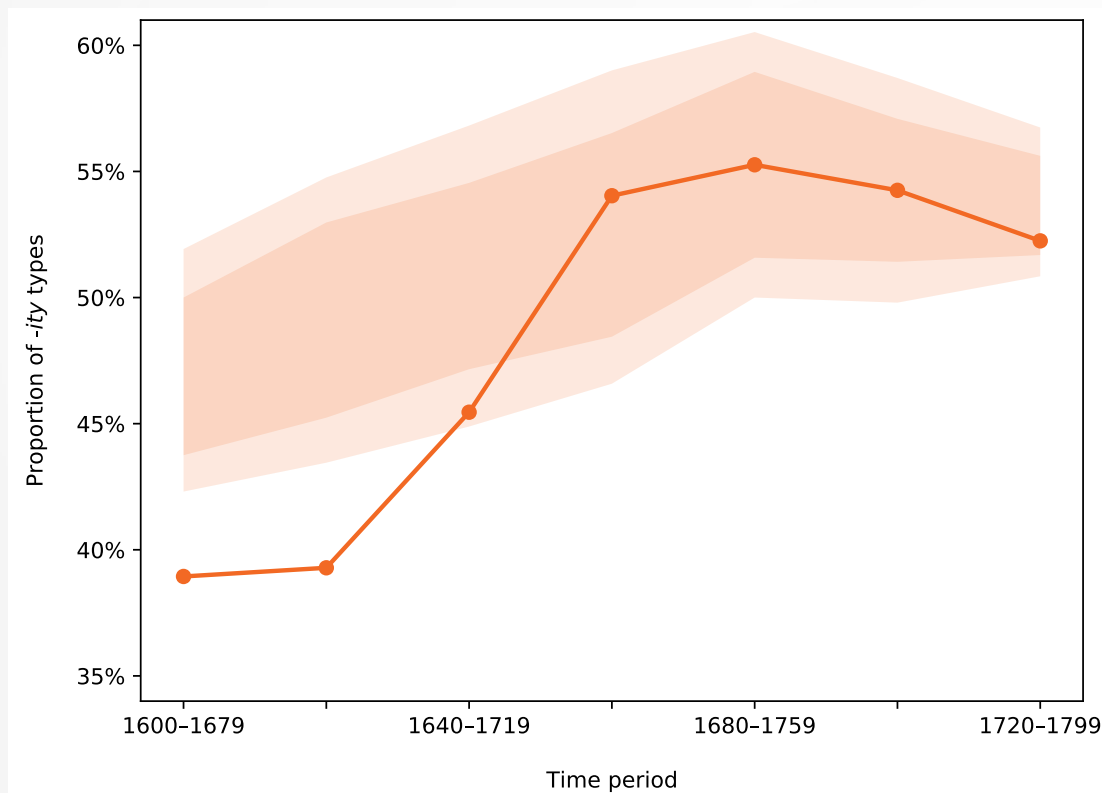
# FOCUS ON A SPECIFIC PERIOD...



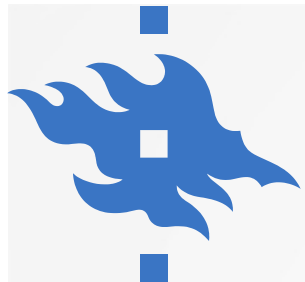
- 1600–1679
- Return corpus size to x-axis
  - Proportion of *-ity* grows nonlinearly with corpus size!
- Compare each gender with random subcorpora of the same size sampled from the entire corpus
  - Proportion of *-ity* statistically significantly **low for women**



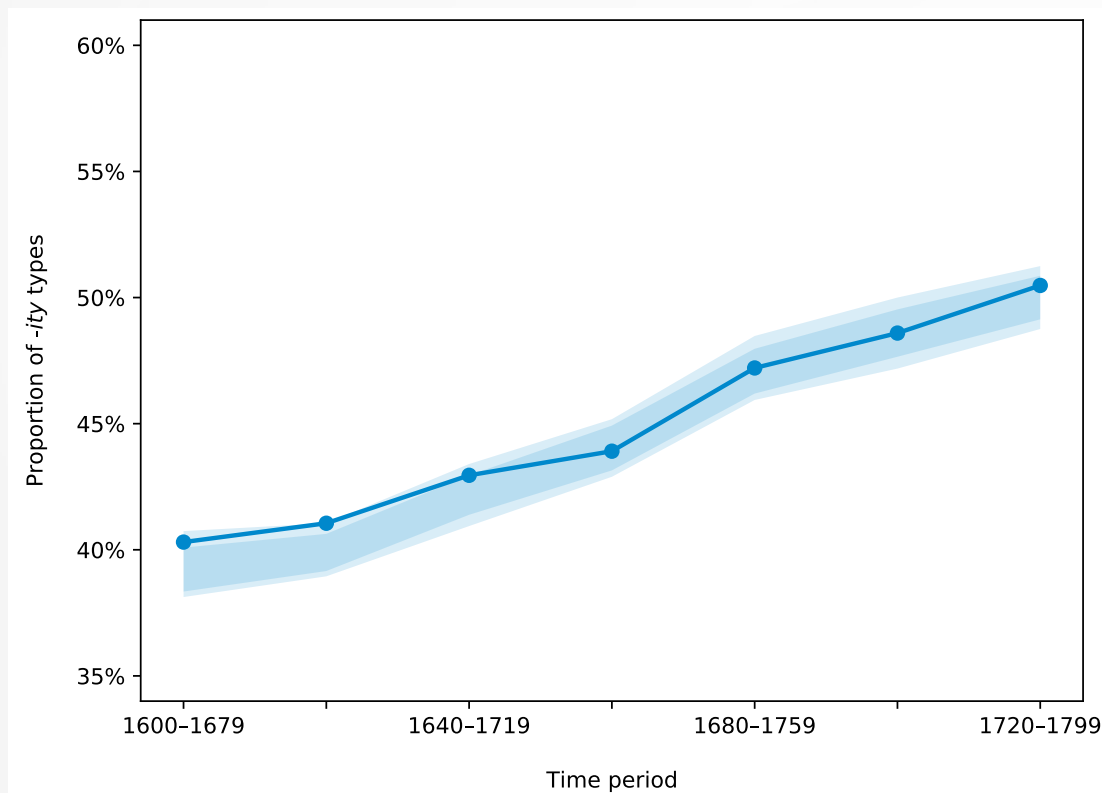
# DO THE SAME FOR ALL PERIODS...



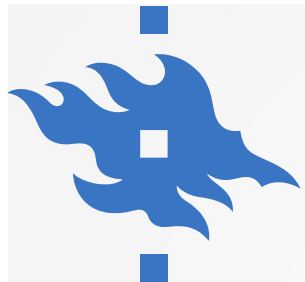
- **Women's** proportion of *-ity* statistically significantly low in the first two periods



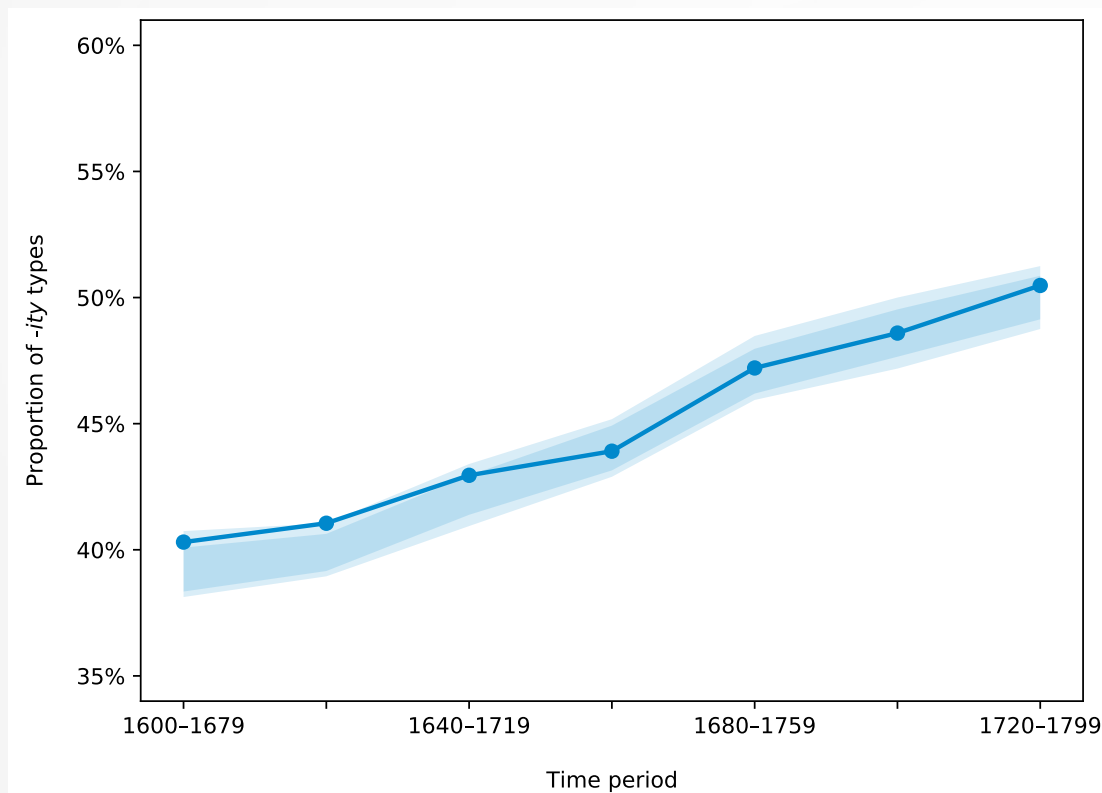
# DO THE SAME FOR ALL PERIODS...



- **Men's** proportion of *-ity* not significantly different from the corpus as a whole in any time period



# DO THE SAME FOR ALL PERIODS...

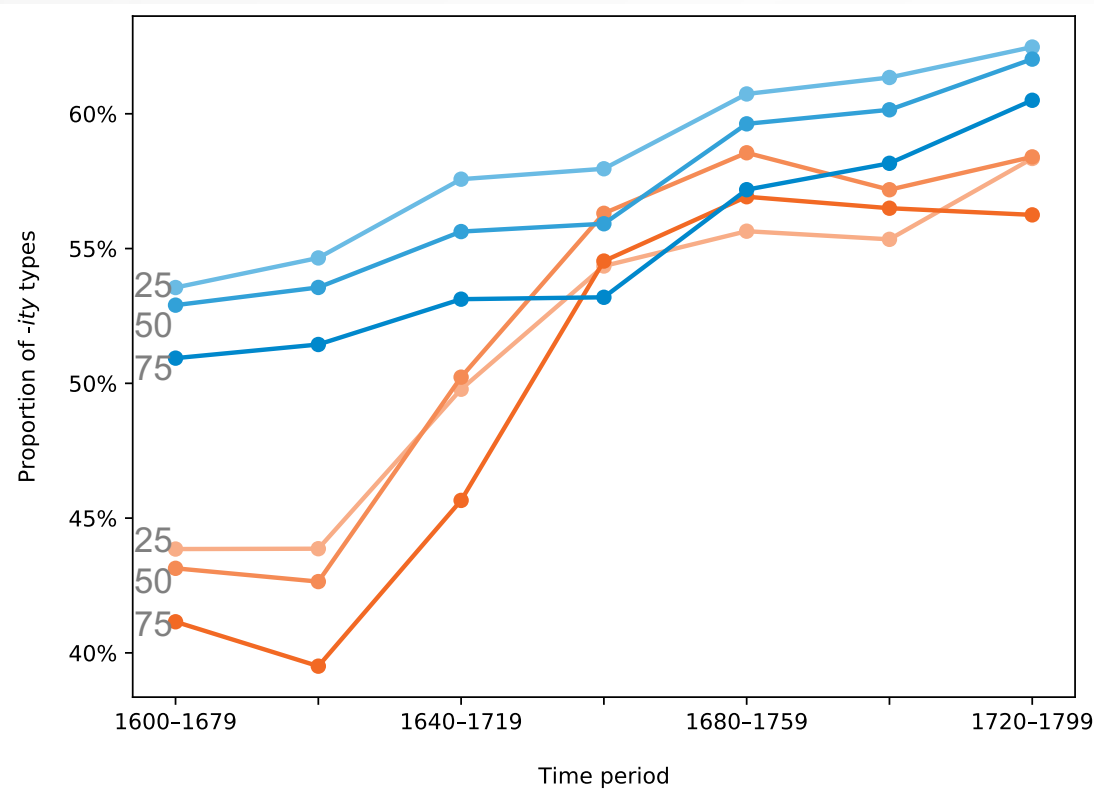


- **Men's** proportion of *-ity* not significantly different from the corpus as a whole in any time period
- **Problem:**
  - Periods not comparable with each other, different amounts of data  
→ we cannot tell whether *-ity* gains ground on *-ness* over time





# TAKE SAMPLES OF EQUAL SIZE FROM GENDER-BASED SUBCORPORA



- 3 corpus sizes: a total of 25/50/75 *-ity/-ness* types
- Proportion of *-ity* increases over time
  - **Men**: steady growth
  - **Women**: lag behind at first, then quickly catch up
    - Statistical significance of lag shown previously

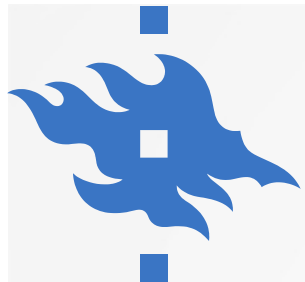


# INTERNAL FACTORS

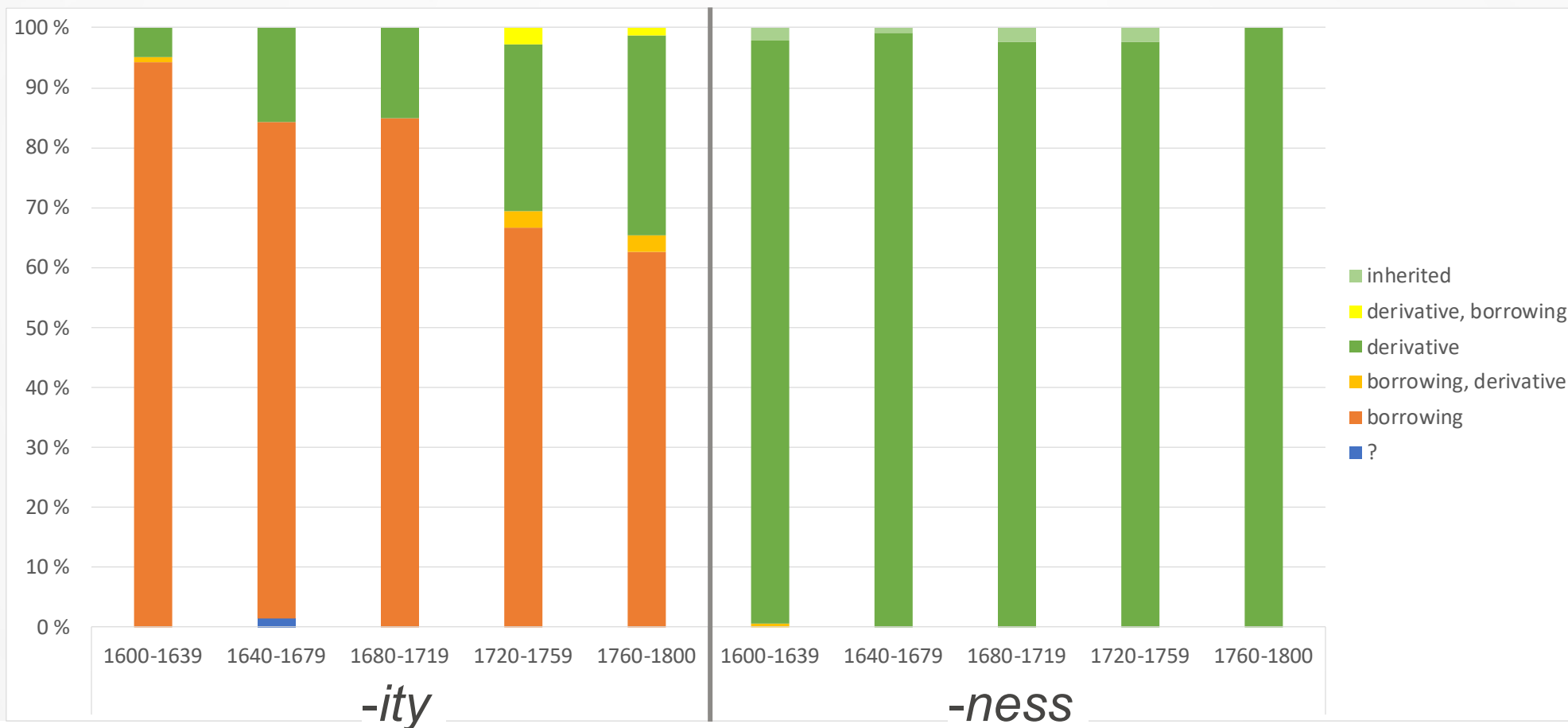


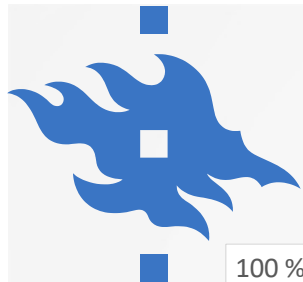
# FACTORS ANALYSED

- **Etymology** (borrowing / derivative); OED
  - e.g. *ability* borrowing, *oddity* derivative
- **Base POS** (usually adjective but others possible as well); OED
  - e.g. *ability*: *able* ADJ, *authorshipness*: *authorship* NOUN
- **Branching structure** (binary / left / right); Hilpert (2013)
  - e.g. [*odd-ity*] binary, [[*un-couth*]-ness] left, [*non*-[*conform-ity*]] right
- **Semantics** (state / thing / office / collectivity); Romaine (1985)
  - e.g. *punctuality*, *electricity*, *officiality*, *laity*
  - Coding slow, only done for a sample of *-ity* instances (derivative)

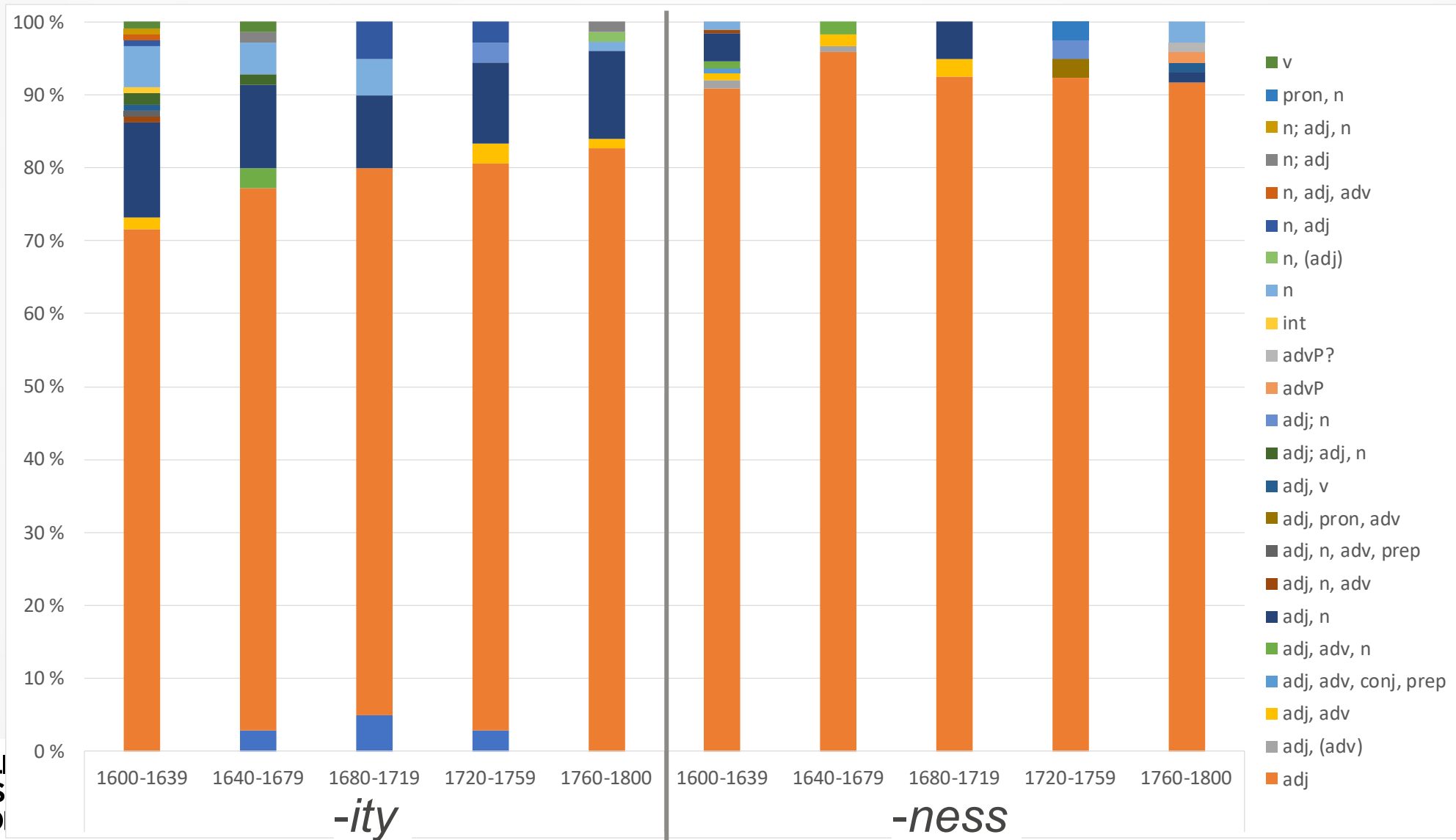


# ETYMOLOGY





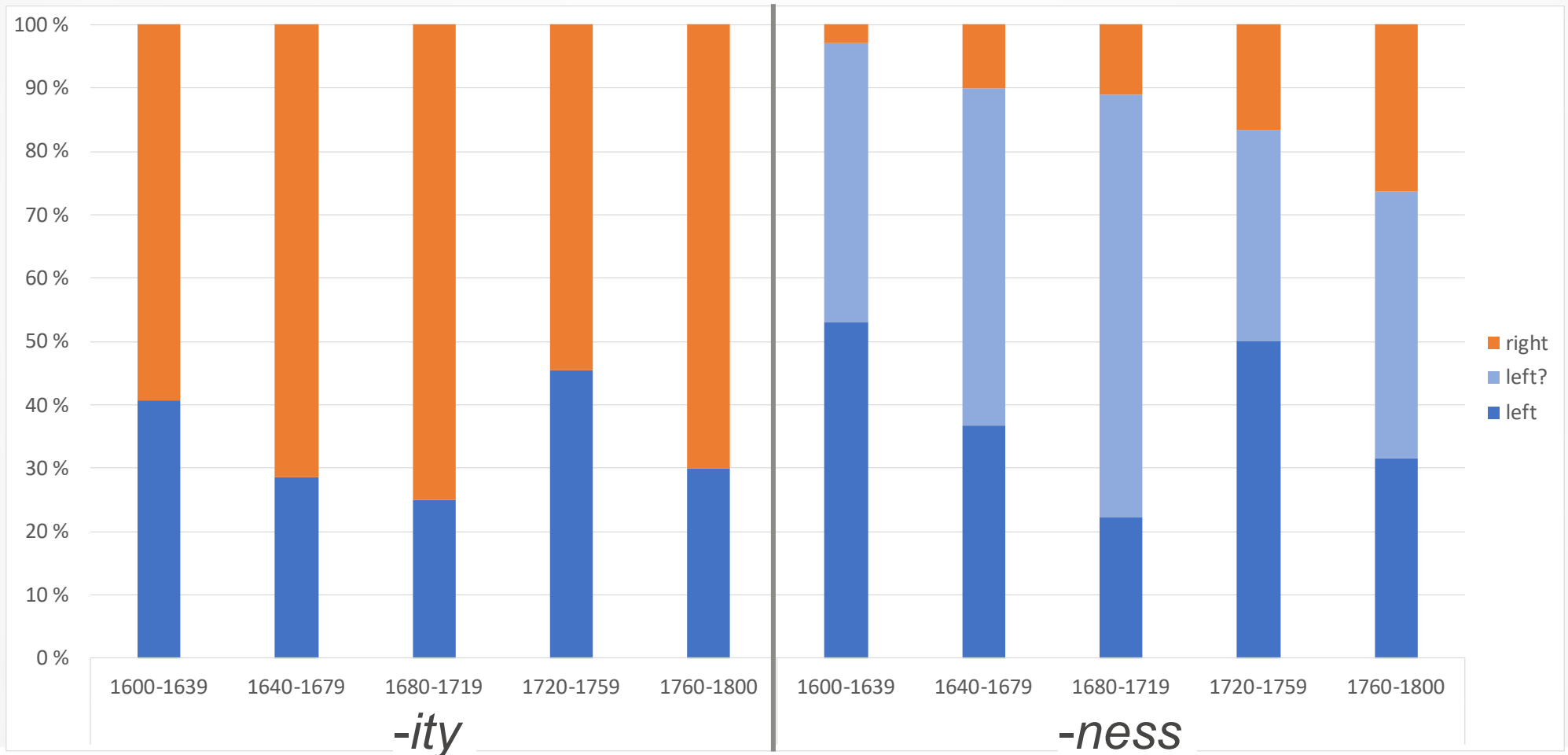
# BASE POS





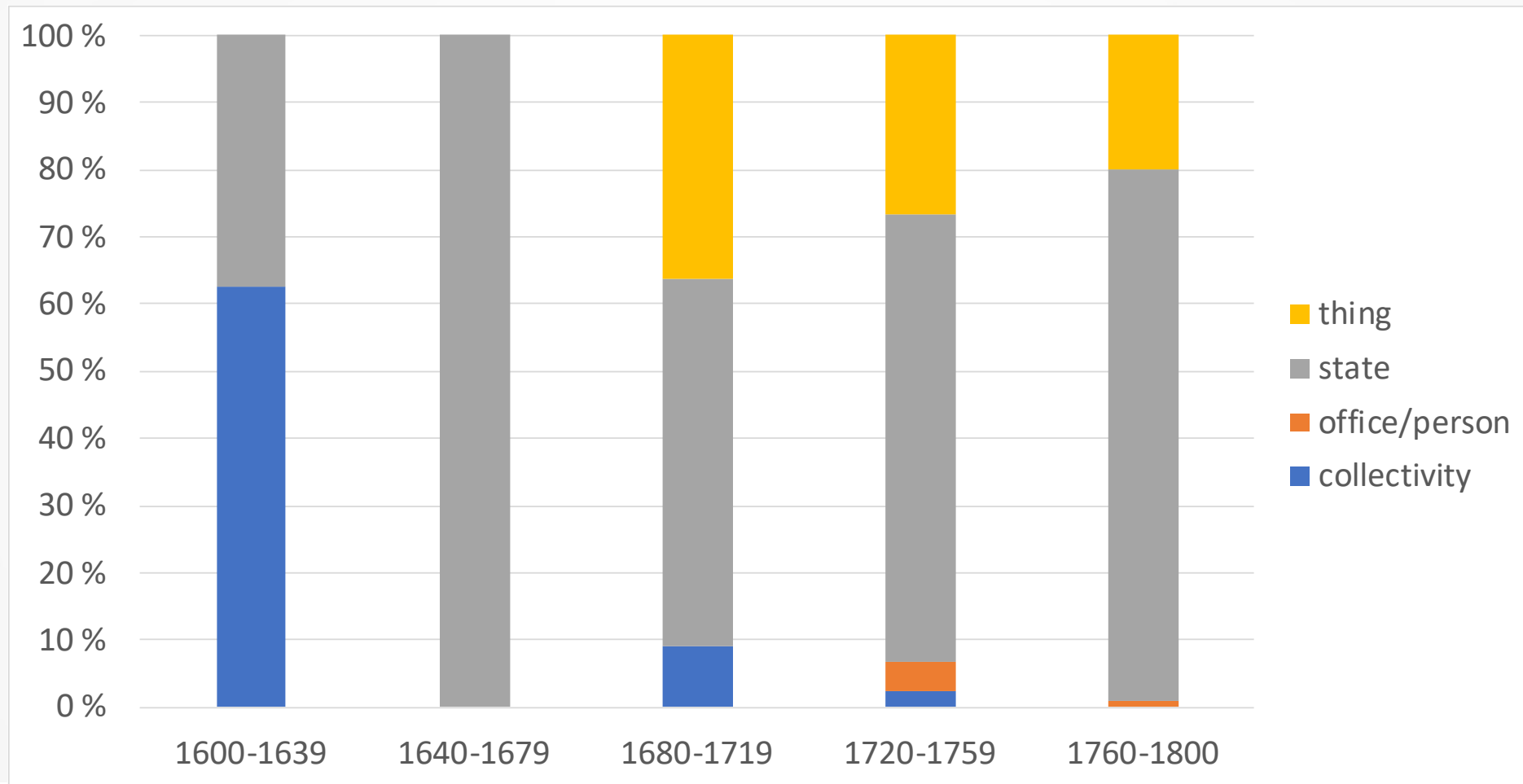


# BRANCHING STRUCTURE

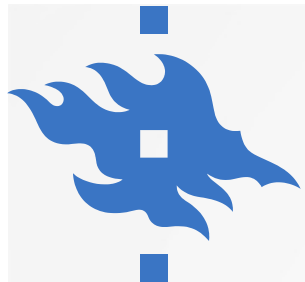




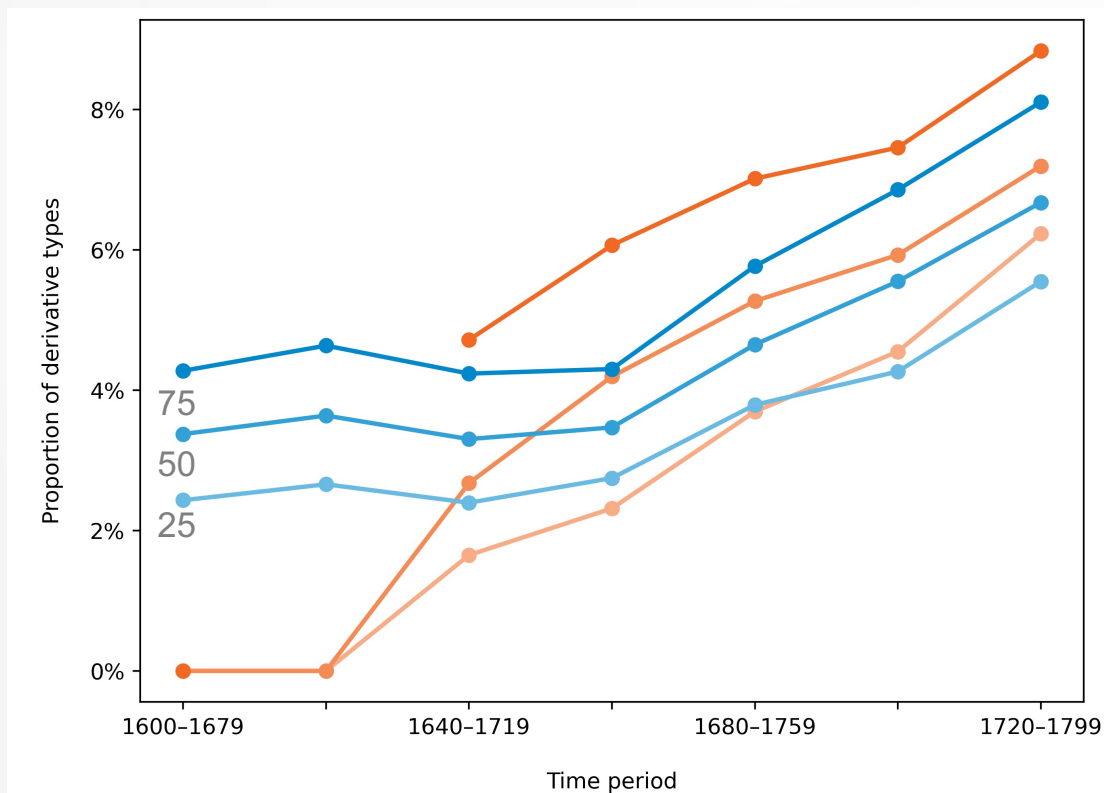
# SEMANTICS (-ITY, SAMPLE OF INSTANCES)



*-ity*



# APPLYING NEW METHODS: -ITY, ETYMOLOGY + GENDER



- Proportion of derived *-ity* types out of all *-ity* types
- Sliding window of 80 years, 20-year increments
- 3 corpus sizes: 25/50/75 *-ity* types
  - **Blue** = men, **orange** = women
- Women lag behind during C17
  - Then quickly catch up with men, and the proportion of derived types only really starts to grow when women join men in using them



# RESULTS

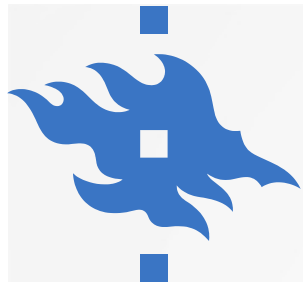
- Etymology
  - *-ity*: share of types derived within English increases over time, women lag behind in C17;  
*-ness*: no change
- Base POS
  - *-ity*: share of adjectival bases increases over time; *-ness*: no clear change
- Branching structure
  - *-ity*: no clear change; *-ness*: share of right-branching, prefixed types increases over time
- Semantics
  - *-ity*: share of ‘state/attribute’ meaning increases over time



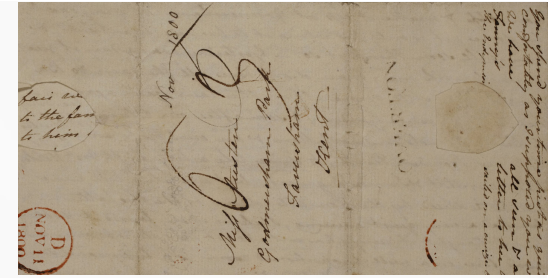
# CONCLUSIONS

- Results support and refine earlier findings
  - **Male-led increase in the productivity of *-ity*** also in relation to *-ness*, more information on diachronic development
- Internal factors, too, point towards increasing productivity of *-ity*
  1. Increase in the share of types originally derived within English
  2. Increase in the share of adjectival bases (types with other bases tend to be borrowed)
  3. Increase in the share of ‘state/attribute’ meaning (others lexicalized)
    - CxG: 2 and 3 surprising – increase in productivity expected to entail use in more contexts, not fewer
- Next step: combined analysis of internal + external factors (cf. 1); *-ness*?





# REFERENCES



- Baayen, R.H. 1993. On frequency, transparency and productivity. G. Booij & J. Van Marle (eds.), *Yearbook of Morphology 1992*, 181–208. Kluwer.
- Bolinger, D. 1948. On defining the morpheme. *Word* 4:18–23.
- Cowie, C. & C. Dalton-Puffer. 2002. Diachronic word-formation and studying changes in productivity over time: Theoretical and methodological considerations. J.E. Díaz Vera (ed.), *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics*, 410–437. Rodopi.
- Hilpert, M. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. CUP.
- Rodríguez-Puente, P. 2020. Register variation in word-formation processes: The development of *-ity* and *-ness* in Early Modern English. *International Journal of English Studies* 20(2):145–167.
- Rodríguez-Puente, P., T. Säily & J. Suomela, submitted. New methods for analysing diachronic suffix competition across registers: How *-ity* gained ground on *-ness* in Early Modern English. *IJCL*.
- Romaine, S. 1985. Variability in word formation patterns and productivity in the history of English. J. Fisiak (ed.), *Papers from the 6th International Conference on Historical Linguistics*, 451–465. Benjamins.
- Säily, T. 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1): 119–141.
- Säily, T. 2014. *Sociolinguistic Variation in English Derivational Productivity: Studies and Methods in Diachronic Corpus Linguistics*. Société Néophilologique.
- Säily, T. & J. Suomela. 2009. Comparing type counts: The case of women, men and *-ity* in early English letters. A. Renouf & A. Kehoe (eds.), *Corpus Linguistics: Refinements and Reassessments*, 87–109. Rodopi.
- Säily, T. & J. Suomela. 2017. *types2*: Exploring word-frequency differences in corpora. T. Hiltunen, J. McVeigh & T. Säily (eds.), *Big and Rich Data in English Corpus Linguistics: Methods and Explorations*. VARIENG. [https://varieng.helsinki.fi/series/volumes/19/saily\\_suomela/](https://varieng.helsinki.fi/series/volumes/19/saily_suomela/)