SOCIOLINGUISTIC VARIATION AND CHANGE IN PRODUCTIVITY

Lecture series on productivity, 18 December 2023

Tanja Säily

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

OUTLINE

INTRODUCTION

CASE 1: *-ITY* & *-NESS*

CASE 2: *VERY/MUCH -ED*

CASE 3: BE *GOING TO* V

CONCLUSION

INTRODUCTION

VARIATION IN PRODUCTIVITY

- Productivity as a **gradient** phenomenon (e.g. Baayen 1992) → focus on variation
  - Initial focus on **morphological productivity** (e.g. *-ity*, *-ness*)
  - Extended to linguistic constructions of all kinds, especially in Construction Grammar (e.g. BE *going to* V)
- **Internal** factors influencing productivity reasonably well studied
  - E.g. semantics: *-ity* is more productive in the sense 'state/quality' (*punctuality*)
- **External**, particularly social factors less so
  - E.g. gender: is *-ity* more productive among men or women?

SOCIOLINGUISTIC VARIATION IN PRODUCTIVITY

- Romaine (1983): sociolinguistic variation in **acceptability judgments** of formations in the nominal suffixes *-ness* and *-ity* (e.g. *productiveness*, *productivity*)

    - Results: variation by **individual, age, gender**

    - Problem: is this related to productivity or to the metalinguistic ability to make judgments of this kind? → other types of evidence needed

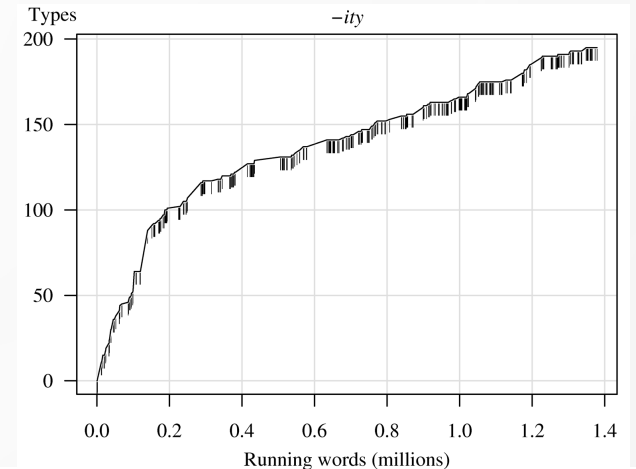- (Two decades pass with nothing much going on, no suitable corpora…)

SOCIOLINGUISTIC VARIATION IN PRODUCTIVITY

- Early **corpus-based** research on sociolinguistic variation and change in morphological productivity

  - E.g. Brezina (2005); Palmer (2009); Keune (2012); Lupica Spagnolo (2013)

- Social factors found relevant: **gender, age, education, region**

  - Keune et al. (2006), *Corpus of Spoken Dutch*: highest affixal productivity is generally exhibited by highly educated older men

- Methodological issue: how to compare productivity measures across subcorpora of different sizes (e.g. men vs. women)?
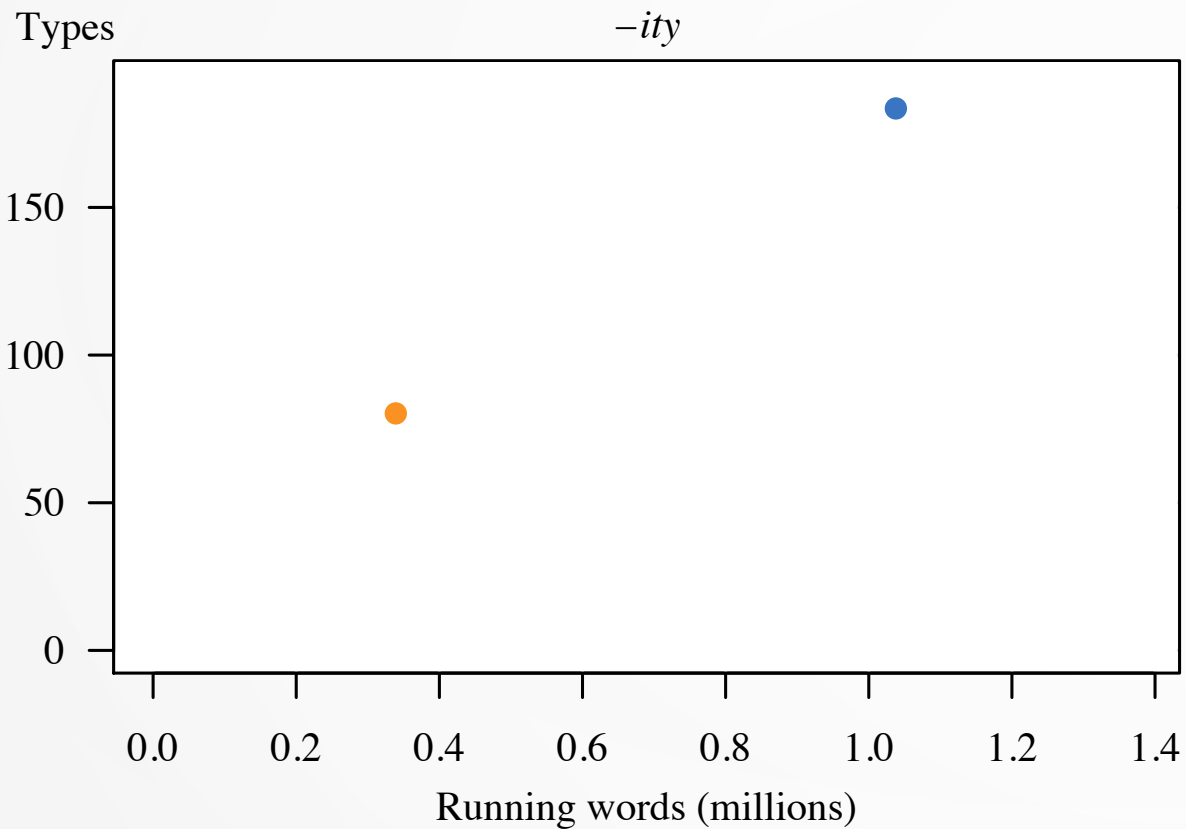
PRODUCTIVITY

- 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 occurring with the element 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
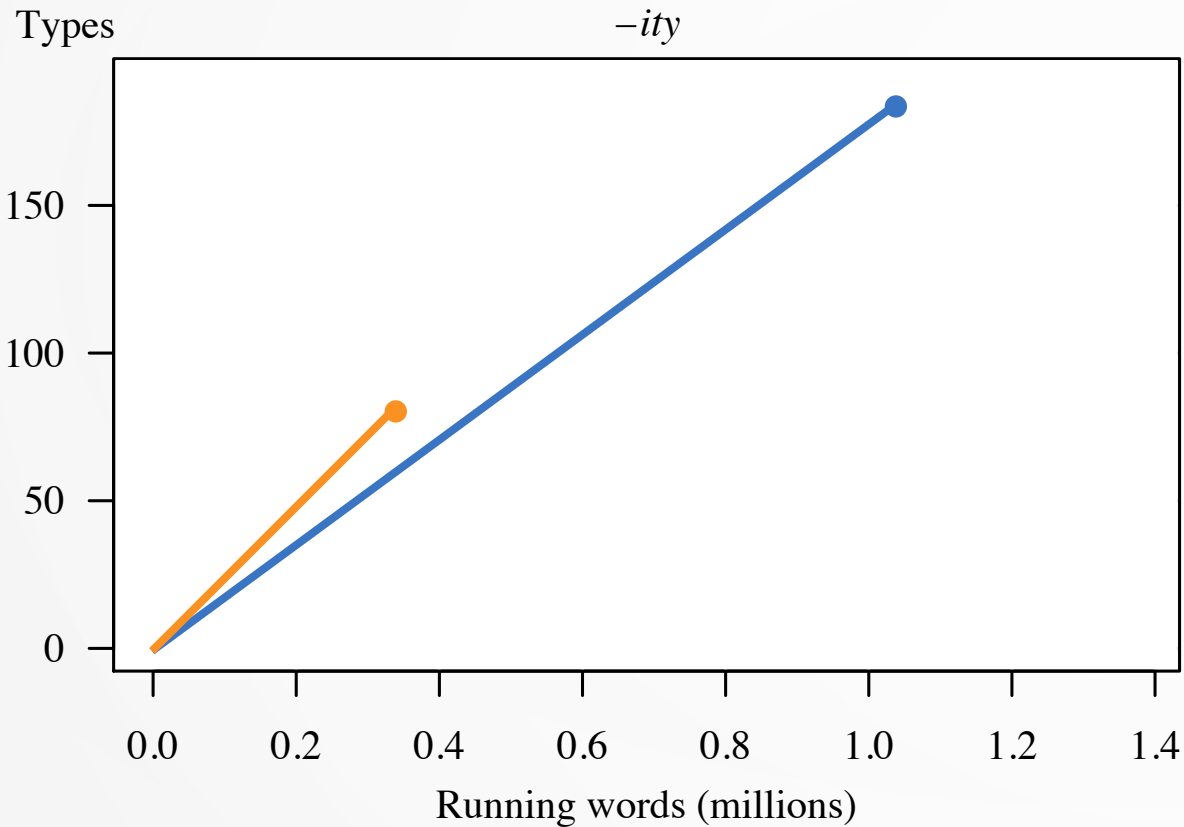   → **normalization not justifiable**

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

NORMALIZATION

Types

−ity

150

100

50

0

0.0   0.2   0.4   0.6   0.8   1.0   1.2   1.4

Running words (millions)

- Who uses comparatively more *-ity* types, **men** or **women**?

NORMALIZATION

Types

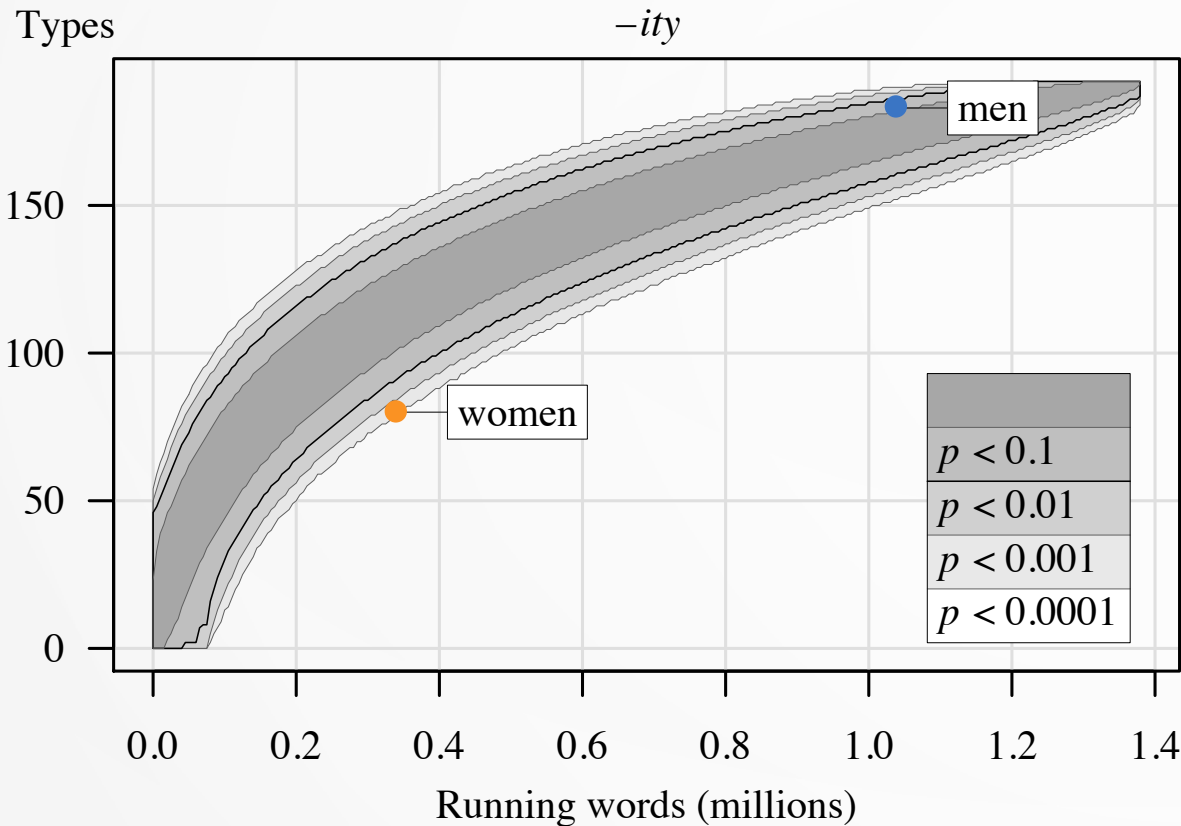*−ity*

- Who uses comparatively more *-ity* types, **men** or **women**?

- Normalization says women, but…

SÄILY & SUOMELA (2009, 2017)

- 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

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
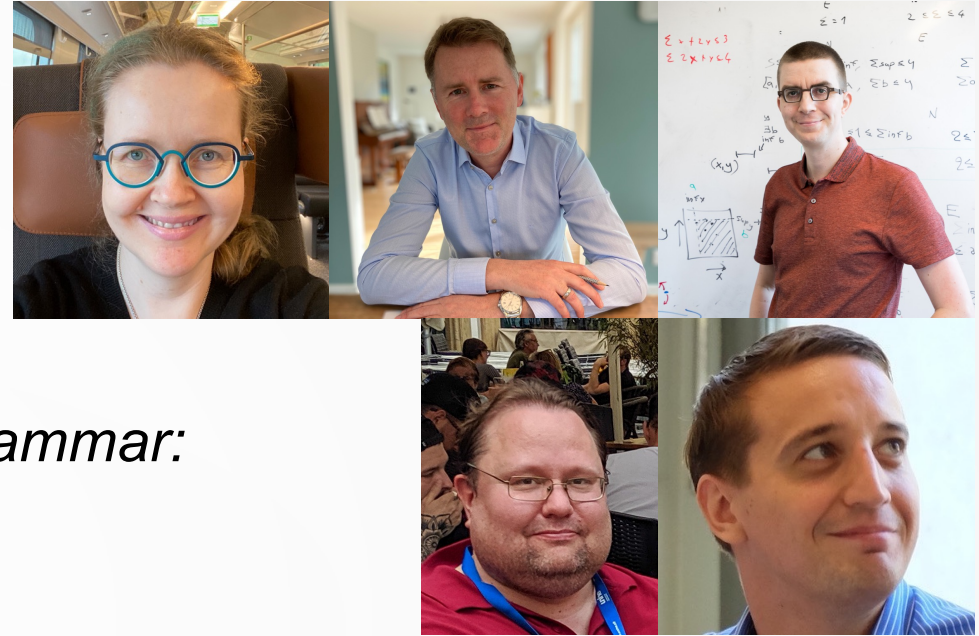
HISCOP PROJECT

- *Historical Sociolinguistics Meets Construction Grammar: The Case of Productivity in English*

  - Research Council of Finland, 2020–2023

  - Funded researcher: **Tanja Säily**

  - Collaborators: Martin Hilpert, Jukka Suomela, Turo Vartiainen, Florent Perek

  - 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

CASE 1: *-ITY* & *-NESS*

Joint work with Martin Hilpert and Jukka Suomela (assisted by Lassi Saario)

-*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. (2022)

-*ITY* AND -*NESS* IN C17–18 PERSONAL LETTERS
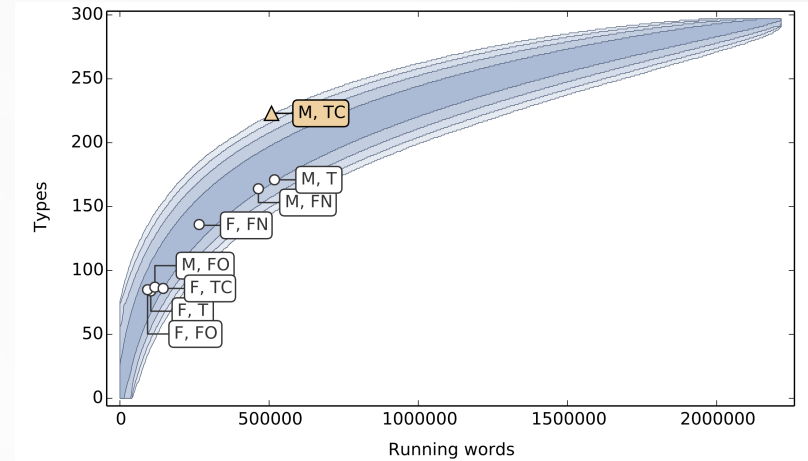
- Säily (2014): **external** factors

  - 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. 2022),
  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 and their interplay with gender

MATERIAL
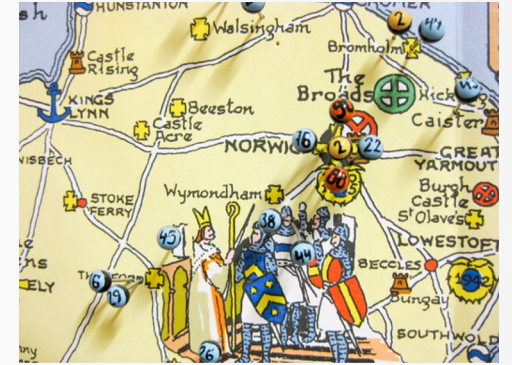
- *Corpora of Early English Correspondence*: personal letters, c. 1400–1800
  - c. 12,000 letters, 1,200 writers, 5 million words (C17–18: 3.5 million words)
  - **Social metadata** on letters, writers, recipients (e.g. gender, social rank)
  - Everyday language use → good for sociolinguistic research

- Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg et al. at the University of Helsinki
  - Based on published editions of letters sampled & digitized by the team

varieng.helsinki.fi/CoRD/corpora/CEEC/

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

SUFFIX COMPETITION

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
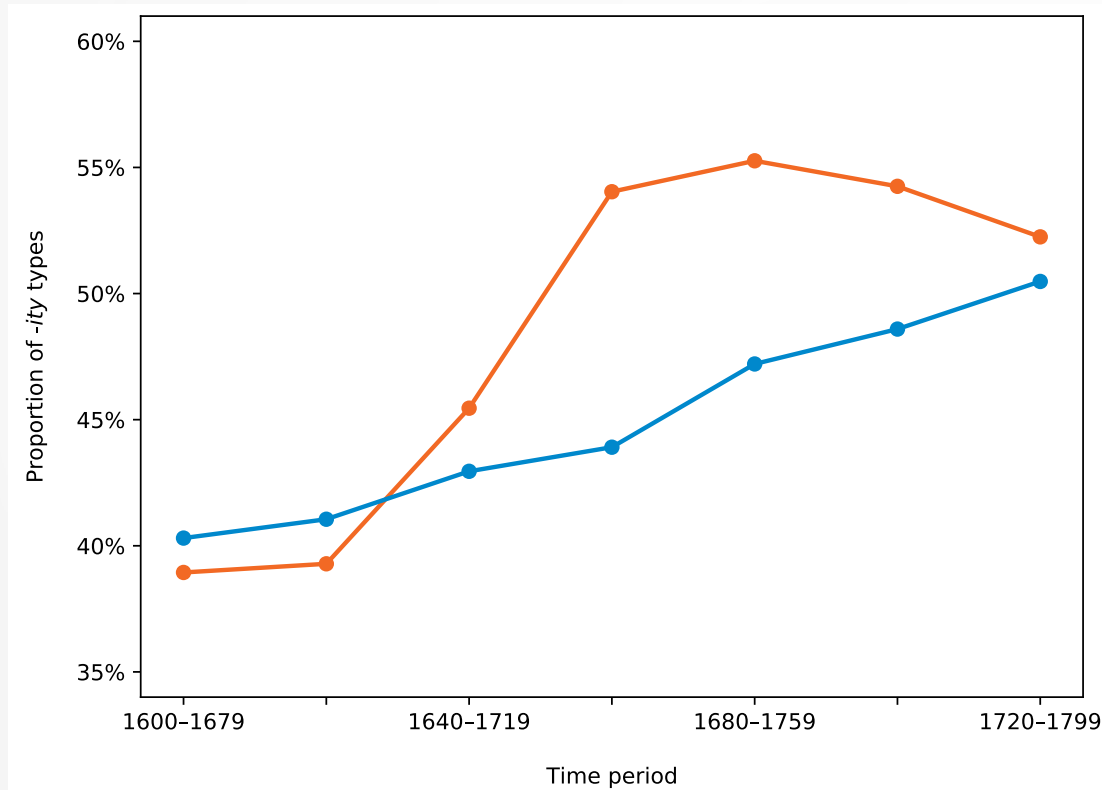UNIVERSITY OF HELSINKI

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

A line chart with "Proportion of -*ity* types" on the y-axis (ranging from 35% to 60%) and "Time period" on the x-axis (ranging from 1600–1679 to 1720–1799). Two lines are shown: a blue line (men) and an orange line (women).

- **Blue** = men, **orange** = women

- 80-year sliding window, 20-year intervals

- **Problems**:

  - Different amounts of data from genders → comparability?

    – Turns out that *proportions of types* grow nonlinearly with corpus size, too! ☹

  - Statistical significance?

TAKE SAMPLES OF EQUAL SIZE FROM GENDER-BASED SUBCORPORA
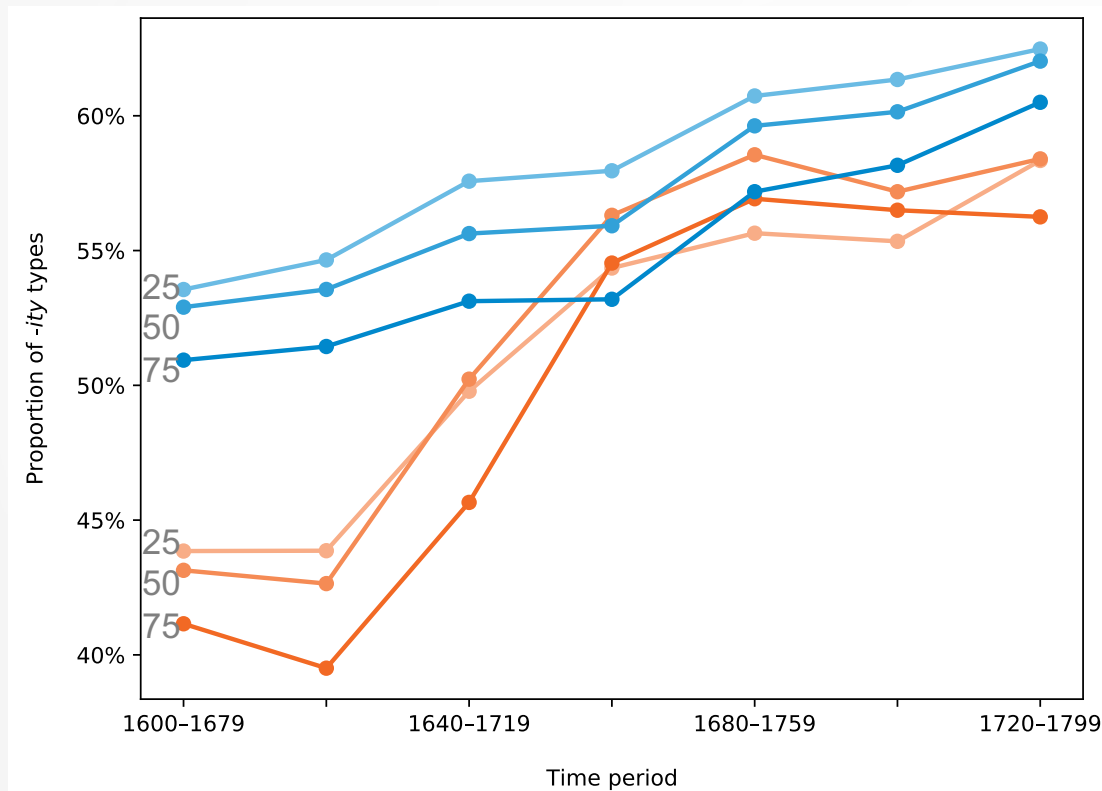
- 3 corpus sizes: a total of 25/50/75 *-ity*/*-ness* types

- Proportion of *-ity* increases over time

  - **Men**: steady growth

  - **Women**: lag behind at first, then quickly catch up

    – But is this statistically significant?

HELSINGIN YLIOPISTO
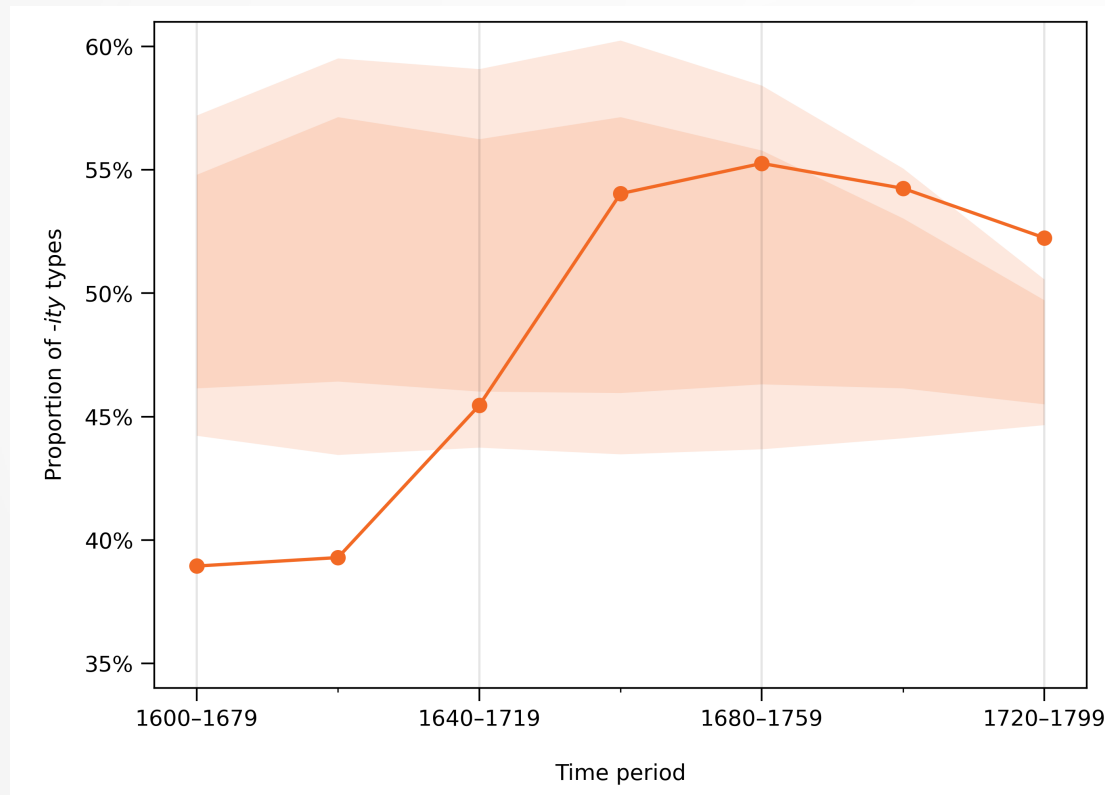HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

SIGNIFICANCE OF GENDER DIFFERENCES
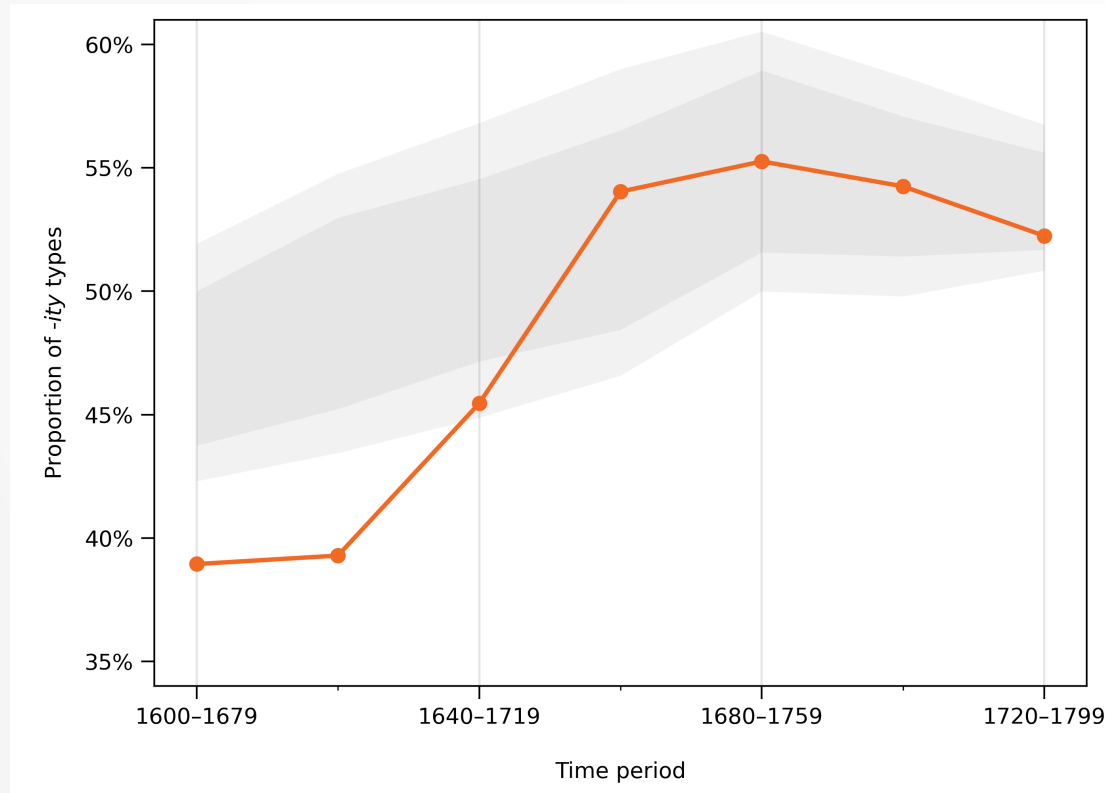
- Compare e.g. women of each period with randomly composed subcorpora of the same period

- Women = orange, random = grey

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

SIGNIFICANCE OF CHANGE OVER TIME

- Compare e.g. women of each period with randomly composed subcorpora of women of all periods

- Women = orange

- Säily et al. (forthcoming)

INTERNAL FACTORS

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

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
- (Not discussed in this presentation: semantic type, occurrence in possessive constructions)
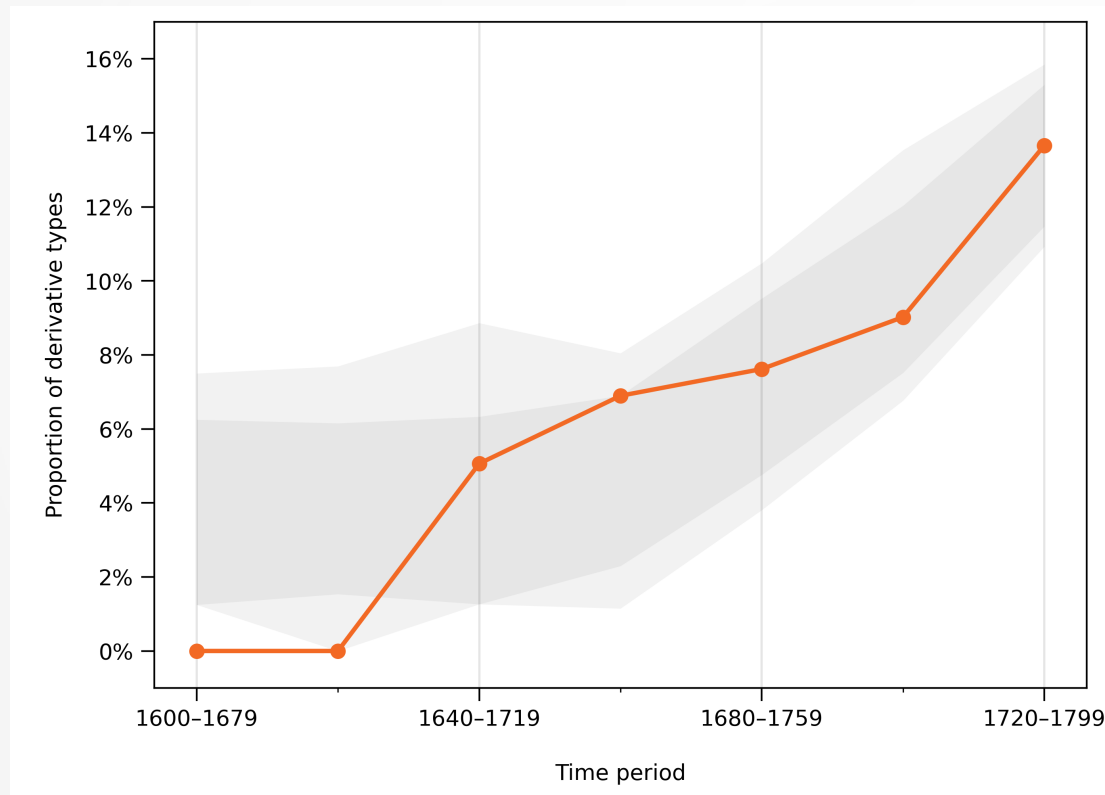
ETYMOLOGY
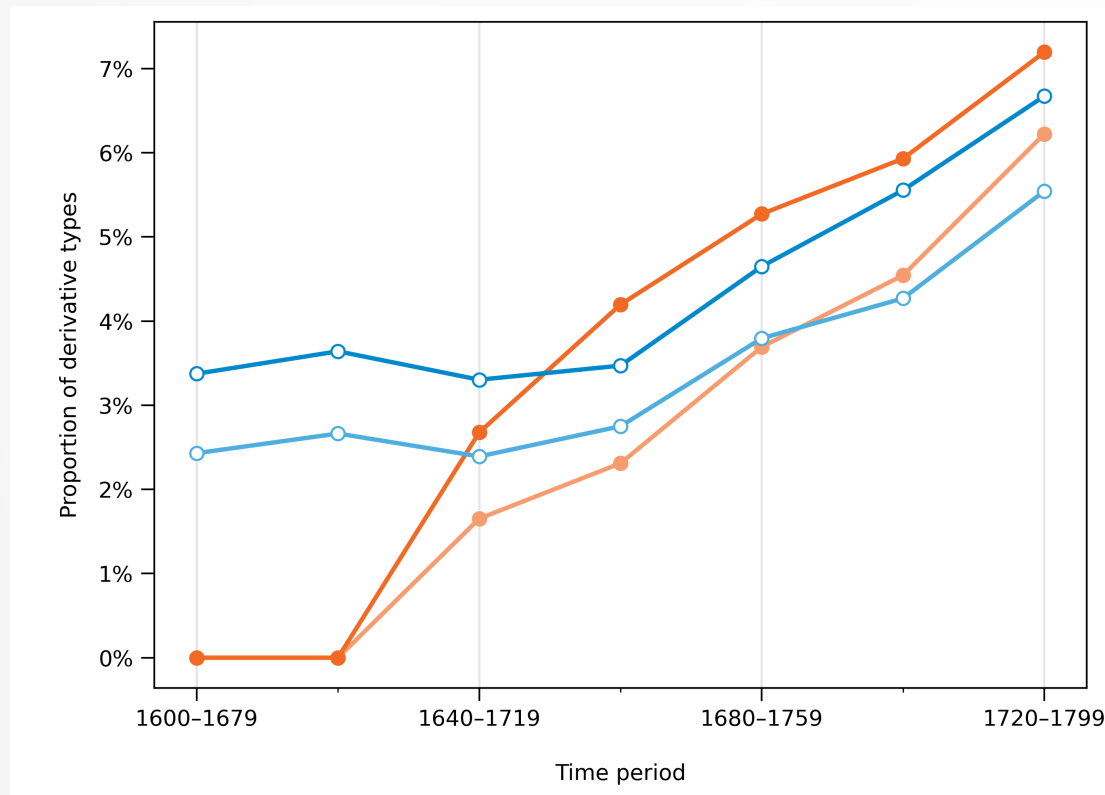
- *-ity*: women lag behind during C17

  - Then quickly catch up with men, and the proportion of **derived** types only really starts to grow when women join men in using them

ETYMOLOGY

- *-ity*: 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

  - 1st period: lag statistically significant ($p < 0.02$)

EXAMPLES

[…] you must needs bee pleased to Visett a place you are soe much concern'd in, and to bee a wittnesse your selfe of the **probabillity** of your hopes […]

(Dorothy Osborne to her future husband, William Temple, 1654)

That space is impenetrable as well as matter, therefore it emplyes something reall or is matter itself. […] And this proves space to have no **impenetrability**.

(Henry More to his friend, Lady Anne Conway, 1651)

EXAMPLES

[…] for she shall never forget your **agreeability** to her in her visits at St Martin's street, during my illness, when she recollects giving you much trouble, which you bore with great patience.

(Frances Burney to her friend, Mrs Francis, 1791)

I protest, it is to me the most difficult of things to write to one of your female geniuses – there is a certain degree of **cleverality** (if I may so call it), […] a […] see-saw of dialogue – which I know no more of than you do of cruelty.
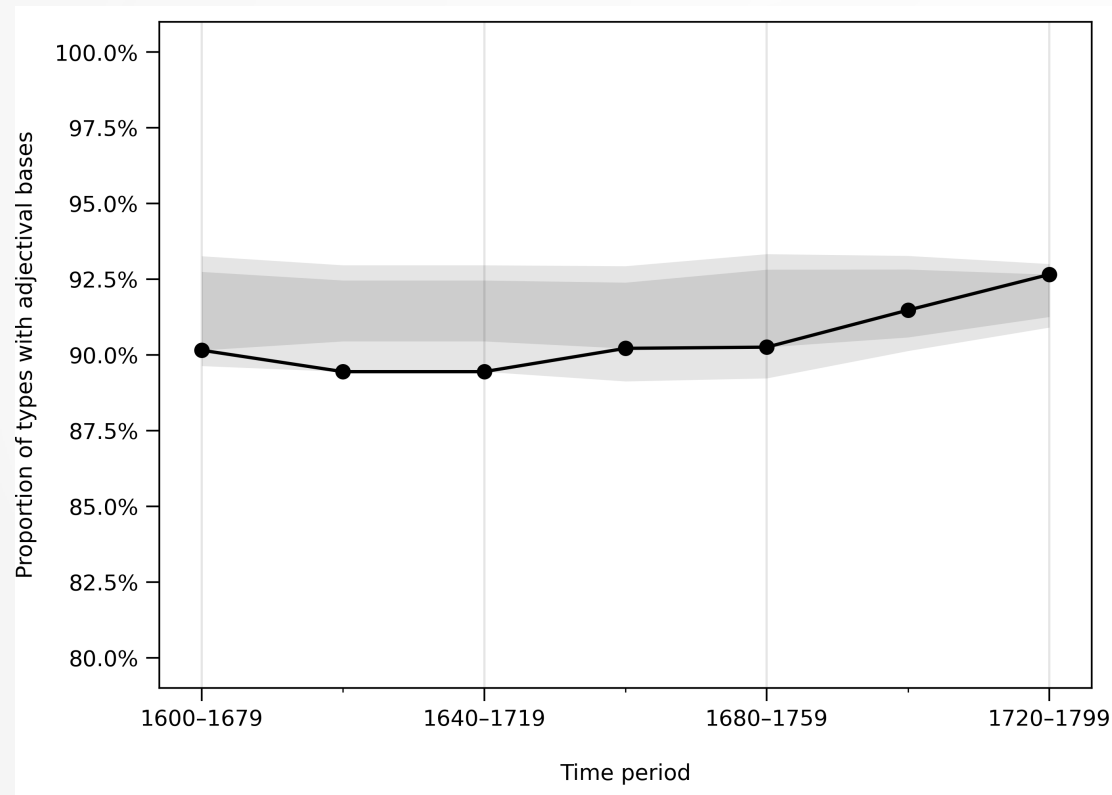
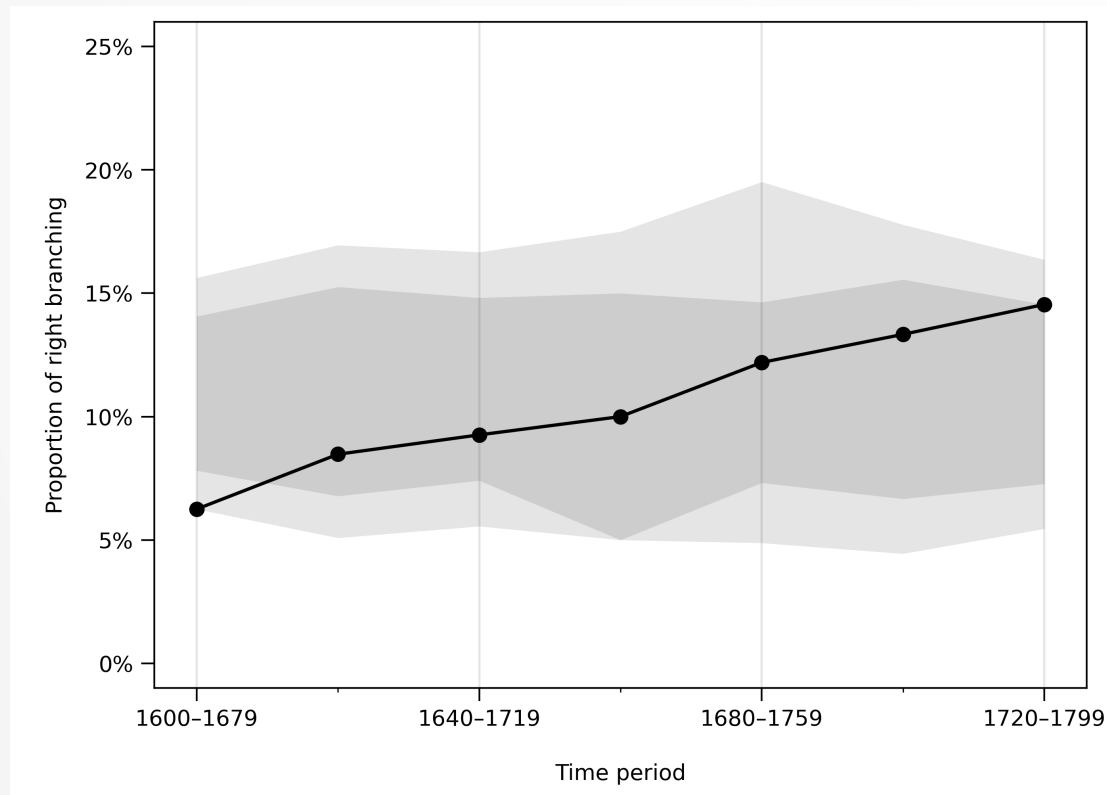(Ignatius Sancho to his friend, Miss Crewe, 1778?)

Wikimedia Commons

BASE POS

- No statistically confirmable trends by gender

- *-ity*: slight increase in share of **adjectival** bases over time

  - Last period: most *-ity* types with non-adjectival bases are earlier borrowings or right-branching

BRANCHING STRUCTURE

- No statistically confirmable trends by gender

- *-ness*: slight increase in share of **right-branching** types over time

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

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

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)

  - CxG: 2 could be related to *coverage*: productivity concentrates on types that are similar to existing types (Suttle & Goldberg 2011)

INTERPRETATION OF SOCIOLINGUISTIC FINDINGS

- Productivity of *-ity*: women lag behind C17, then quickly catch up with men – why?

- *-ity* spreads from more **formal written registers** to more speech-related ones (Rodríguez-Puente 2020; Rodríguez-Puente et al. 2022)

  - Women did not have as much access to formal written registers as men

- *-ity* is a **borrowed, learned and prestigious** suffix

  - Initially easier to use for men with a classical education

  - Better match to men's more informational style of letter-writing (Säily et al. 2017)

- C18: development of a **shared, polite style** among the middle + upper classes

  - Linked to changing conceptions of gender, advances in women's education? (Säily 2014)

CASE 2: *VERY/MUCH -ED*

Joint work with Turo Vartiainen (and Jukka Suomela)

INTRODUCTION

Study based on 2 Research Council of Finland funded **postdoctoral projects**:

- *Categorization, creativity and change in Construction Grammar*

  - Turo Vartiainen

- *Historical sociolinguistics meets Construction Grammar: The case of productivity in English*

  - Tanja Säily

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

CATEGORY CHANGE

- A change where a word from one word class starts to be used in contexts associated with words of another class

- Often accompanied with a change in meaning, but not necessarily

- Our focus: change of **-*ed* participles from verbs to adjectives**

  - e.g. Denison (1998), Vartiainen (2016, 2021)

  - Ambiguity between the **passive** construction (verbal) and the **predicative** construction (adjectival)

VERB OR ADJECTIVE?

1. Miss Day was **frightened** by the cats last night. (COHA, Fiction, 1888)

2. He has been **much interested** in your movements. (COHA, Fiction, 1846)

3. We'd be **very much pleased**. (COHA, Fiction, 1909)

4. We are **very pleased** with the court's ruling. (COHA, News, 2017)

CHANGE IN DEGREE MODIFICATION PATTERNS OF *-ED* PARTICIPLES

- Early C19:   *She's **much** interested in it.*

→ PDE: *She's **very** interested in it.*

- Previous research: gradual change between mid-C19 and mid-C20
    - e.g. Denison (1998), Vartiainen (2021)
    - **Limited** datasets, many grammatical and sociolinguistic aspects remain to be explored
    - Analyses focused on **token frequency** rather than type frequency (productivity; Baayen 2009, Säily 2014, Säily et al. 2018)

OUR STUDY

- Material: *Corpus of Historical American English* (COHA), fiction section (200Mw)
  - Enriched with gender metadata by Öhman et al. (2019)
- **Research questions**
  - Timing of the increase in productivity of *very -ed*?
  - Role of **women vs. men** in the change?
    – Cf. Nevalainen & Raumolin-Brunberg (2003), Säily (2014)
  - *Very much -ed* as a bridging context?
  - (Not discussed in this presentation: role of *-ed* participles denoting psychological states?)

DATA RETRIEVAL

- Retrieved from Korp, https://korp.csc.fi

- **Basic principle 1**: the *-ed* form must be a participle, i.e. a verb form

  - Potential verbal origins checked in OED Online

  - Dozens of *-ed* adjectives excluded from the search results
  (e.g. *warm-hearted*, *hairy-faced*, *good-natured*)

  - Participles with the prefix *un-* only included if the OED includes a verb with the prefix
  (e.g. *unsettle > unsettled*)

- **Basic principle 2**: *much* and *very* must be in variation in the given context

  - *Too much -ed*, *as much -ed* excluded

  - *So* included, even though the meanings of *so much* and *so very* are not exactly the same

METHODS
RODRÍGUEZ-PUENTE ET AL. (2022), SÄILY & SUOMELA (2009, 2017), SÄILY & VARTIAINEN (FORTHCOMING)

- Problem: type frequencies cannot be normalized, grow nonlinearly with corpus size

- **Basic measure**: proportion of types of interest out of all relevant types

  - E.g. proportion of *very -ed* types out of *very -ed* and *much -ed* types

- **Setting 1**: make subcorpora comparable by taking samples of equal size from each

  - Plot proportions as line graphs over time, use a sliding window for periodization

- **Setting 2**: estimate statistical significance of differences observed

  - Use permutation testing within each time period, compare each subcorpus with randomly composed subcorpora of the same size

  - Plot the subcorpus as a line graph over time and the random subcorpora as confidence intervals around the line, use a sliding window for periodization
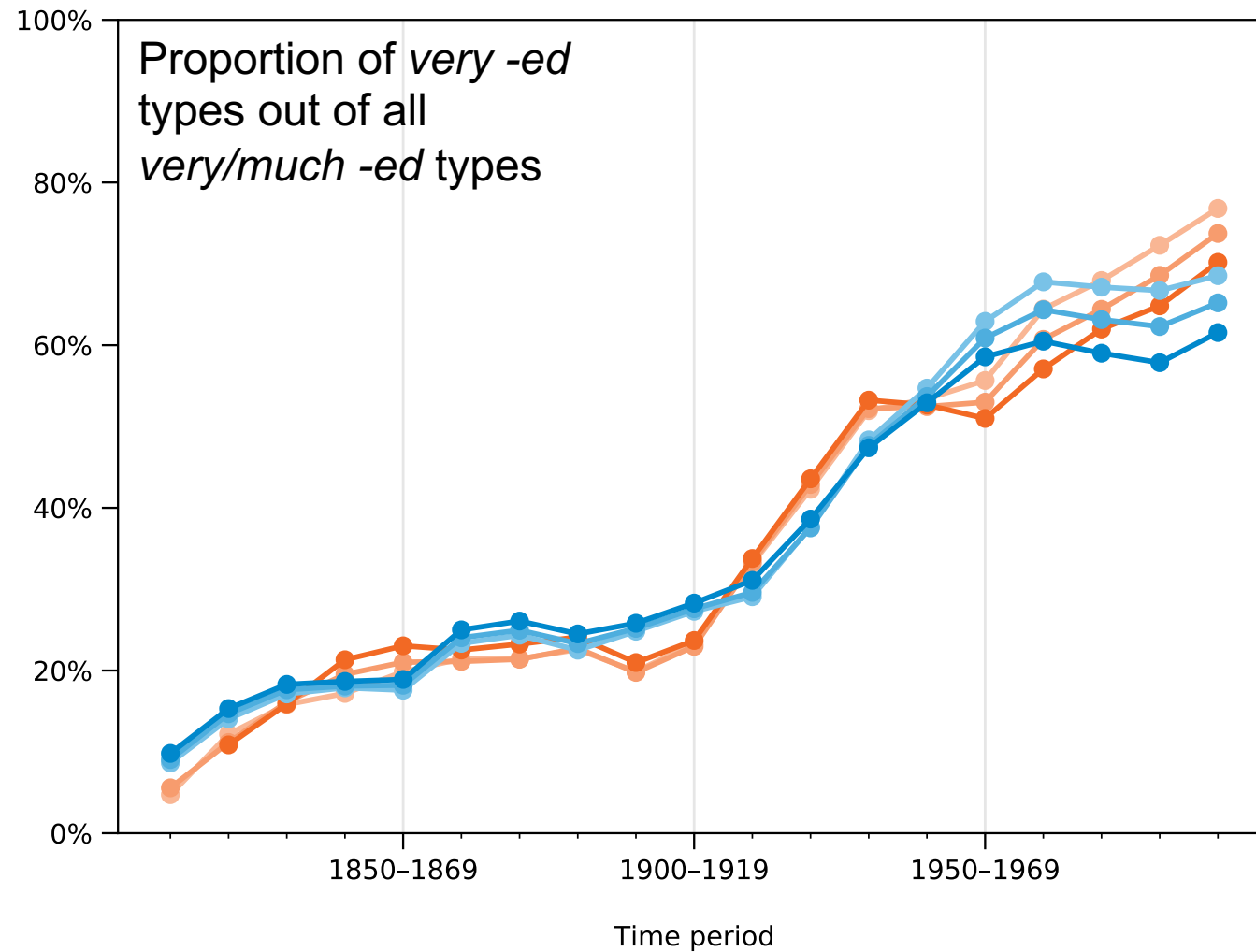
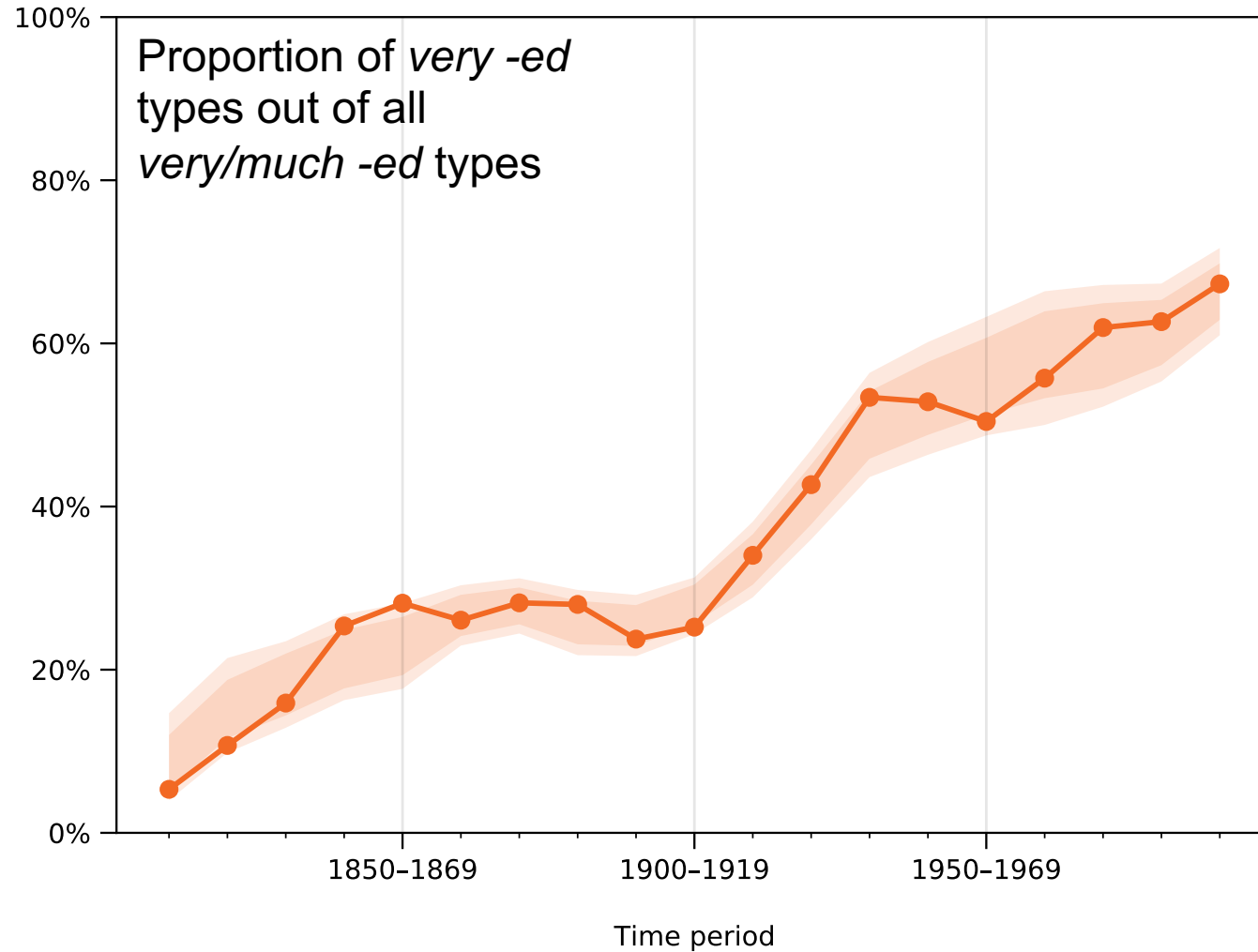# OVERALL DEVELOPMENT OF *VERY -ED*
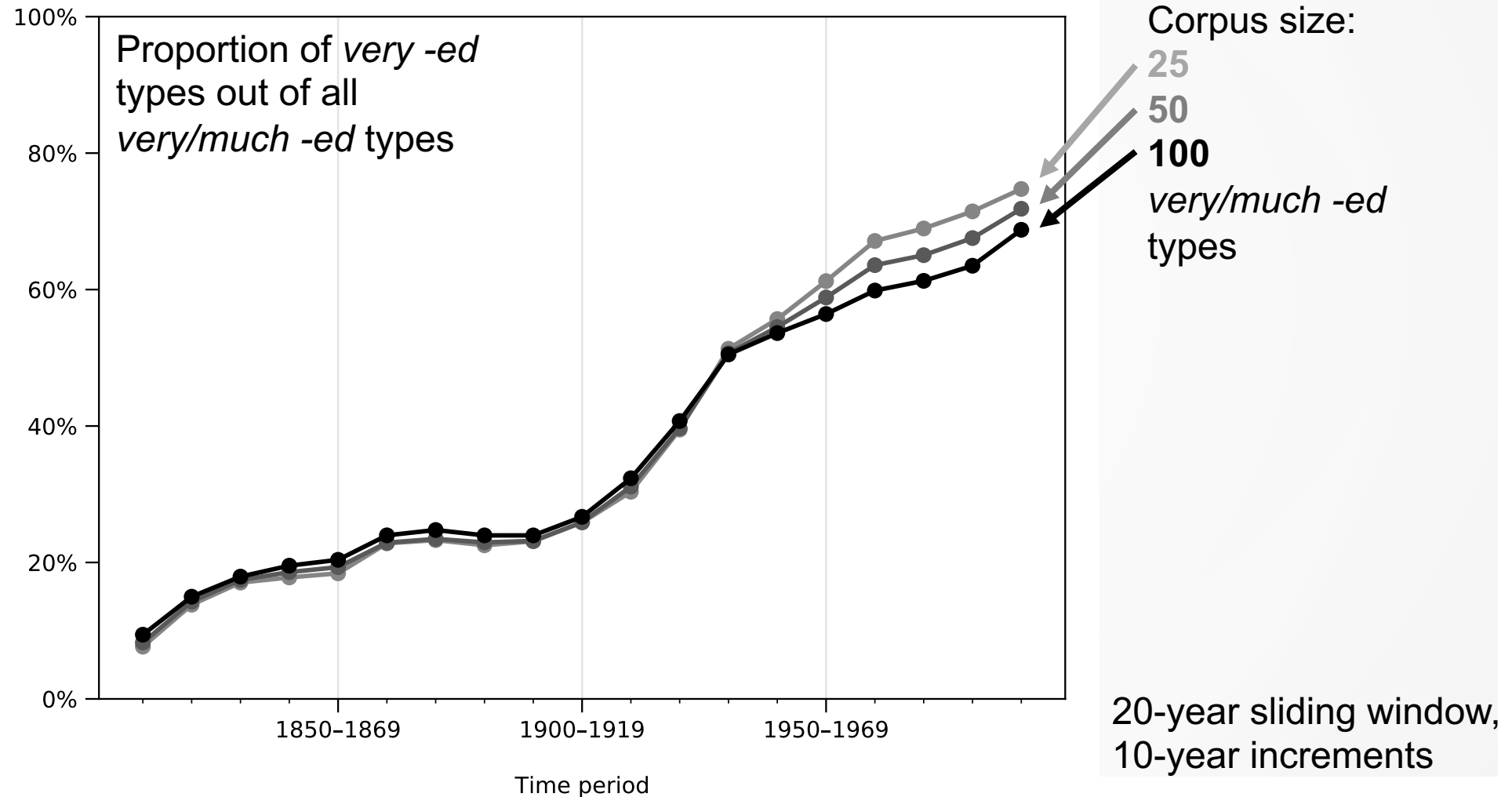
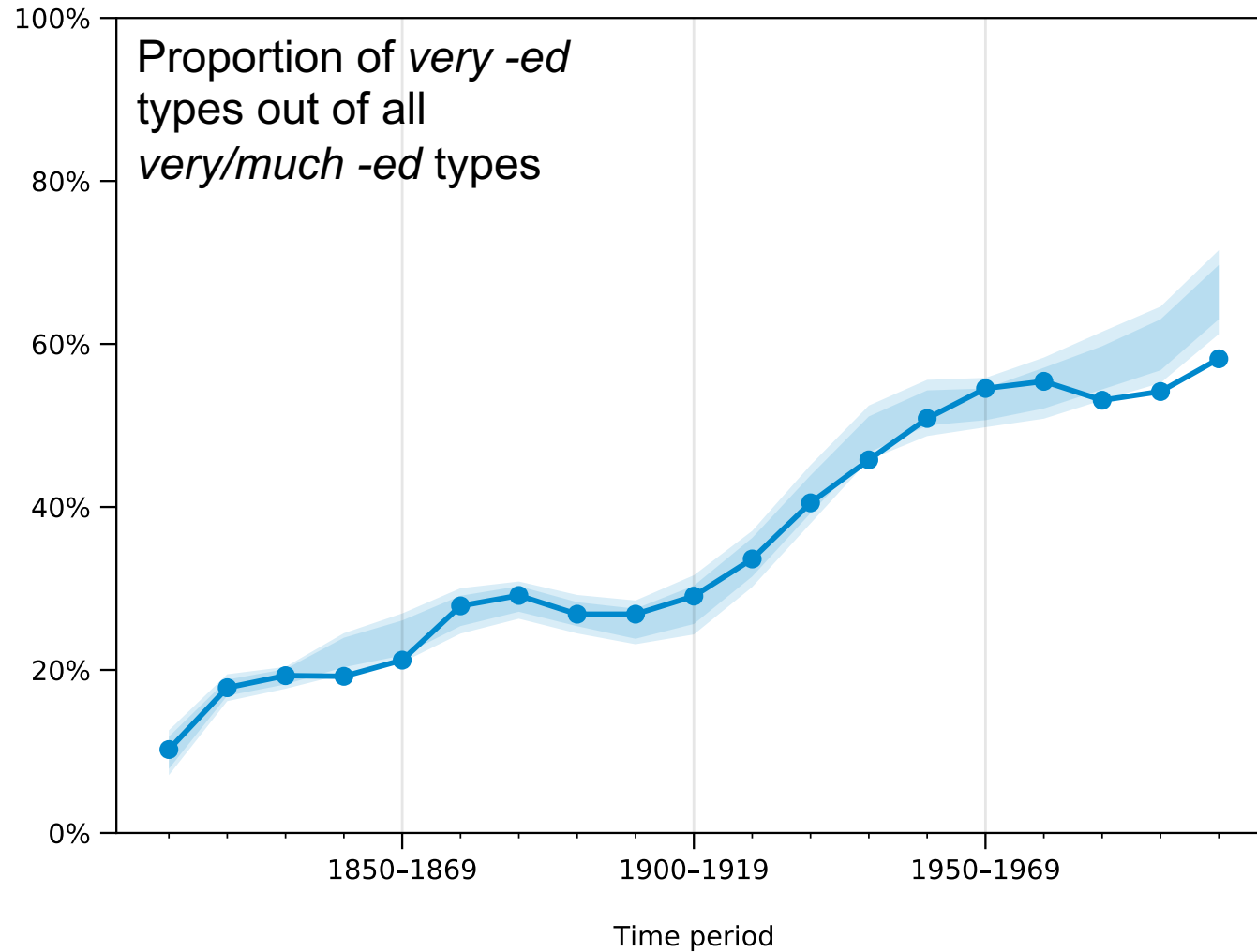Proportion of *very -ed* types out of all *very/much -ed* types

Corpus size:

25

50

100

very/much -ed types

20-year sliding window, 10-year increments

WOMEN (ORANGE) VS. MEN (BLUE)

Proportion of *very -ed*
types out of all
very/much -ed types

STATISTICAL SIGNIFICANCE (WOMEN)

Proportion of *very -ed* types out of all *very/much -ed* types

Time period

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

STATISTICAL SIGNIFICANCE (MEN)

Proportion of *very -ed* types out of all *very/much -ed* types

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

STATISTICAL SIGNIFICANCE (MEN, NOVELS)

Proportion of *very -ed* types out of all *very/much -ed* types

Time period

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

VERY MUCH -ED AS A BRIDGING CONTEXT?

Proportion of *very much -ed* types out of all *much -ed* types

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
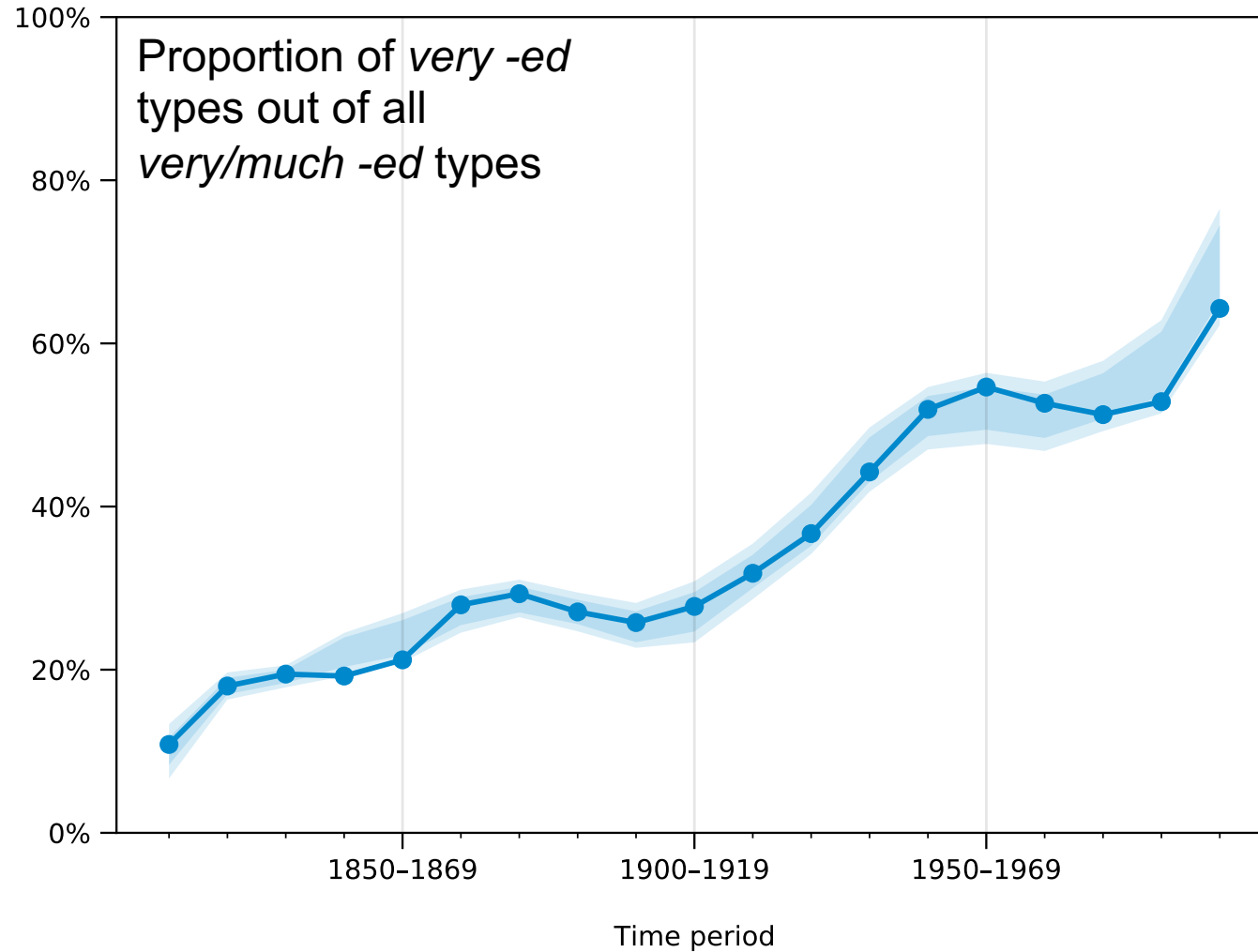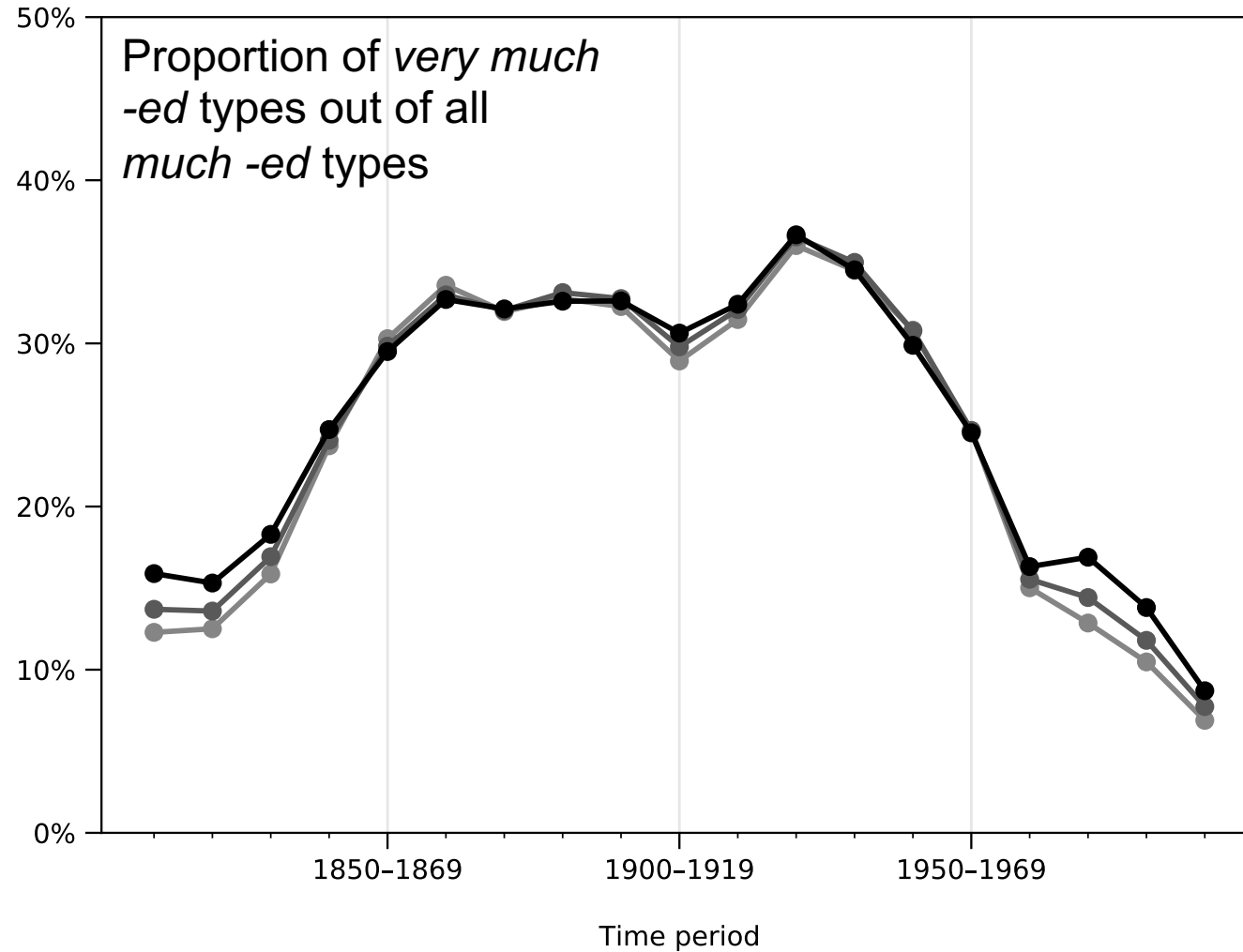
SUMMARY

- *Very -ed* gains ground on *much -ed* in AmE fiction, 1810–2009

  - Change in productivity resembles an **S-curve**: slow beginning, rapid middle starting from beginning of C20, slower final decades of the corpus

- No statistically significant **gender differences** when genre imbalance accounted for (novels vs. others)

- *Very much -ed* could have functioned as a **bridging context**

  - High proportion out of all *much -ed* types overlaps temporally with increase in *very -ed*

CONCLUSION

- **Future research**
    - Attributive vs. predicative usage (*very* initially used more with attribution)
    - Complementation patterns (*by*-phrases, PP-complements, clausal complements)
    - Modifiers of *much* (C19: often modified by *so/very*, PDE: less frequent)
    - Negation (*much* more associated with negation than *very*)
    - Multivariate analysis with grammatical and social factors?
- **Enriched datasets**: great potential for research
    - We can now do sociolinguistics with COHA!

CASE 3: BE *GOING TO* V

Joint work with Florent Perek and Jukka Suomela

GRAMMATICALIZATION OF BE *GOING TO* V

1. *I'm going to the market to buy bananas*   'motion with intention'
2. *I'm going to read your work tomorrow*   'motionless intention'; EModE
3. *There's going to be some serious trouble here* 'prediction'; LModE–PDE
   a. *You're going to **feel** very foolish* (mental verb; COHA, 1932)
   b. ***It's** going to rain* (inanimate subject, it; COHA, 1811)
   c. *Father Paul was going to **be cheated** of his share* (passive voice; COHA, 1946)

(Budts & Petré 2016; Wu et al. 2016)

RESEARCH QUESTIONS

1. How is the grammaticalization reflected in the **productivity** of the construction in LModE–PDE?

    • Internal factors: semantics of the verb (mental verbs), inanimate subject (*it*), passive voice

2. Did the social factor of **gender** play a role in the process?

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MATERIAL

- *Corpus of Historical American English* (COHA)
  - 400 Mw, 1810–2009

- Fiction section: c. 50% of the data
  - **Gender metadata** for authors developed by Öhman et al. (2019)
  - Promising material for sociolinguistic investigation: a more speech-like genre (dialogue)
  - Types of fiction (e.g. short stories, drama, movie scripts) unevenly distributed over time (Säily & Vartiainen forthcoming)
    → restriction to **novels only, c. 150 Mw**

- List of mental verbs from Halliday & Matthiessen (2014: 256–257)

METHODS

- We study the productivity of BE *going to* V by studying **type frequencies**

    - I.e. how many different verbs follow BE *going to*
      in different time periods

- Key challenges:

    - Different amounts of text from different time periods,
      different amounts of text from men and women:
      how to **compare** type frequencies?

    - If we observe trends, are they **statistically significant**?

Types in subcorpora with 18811353 words

Choose **random subcorpora** with the same number of words from each time period

See what the average number of types is in such corpora

novels, be-going-to-verb
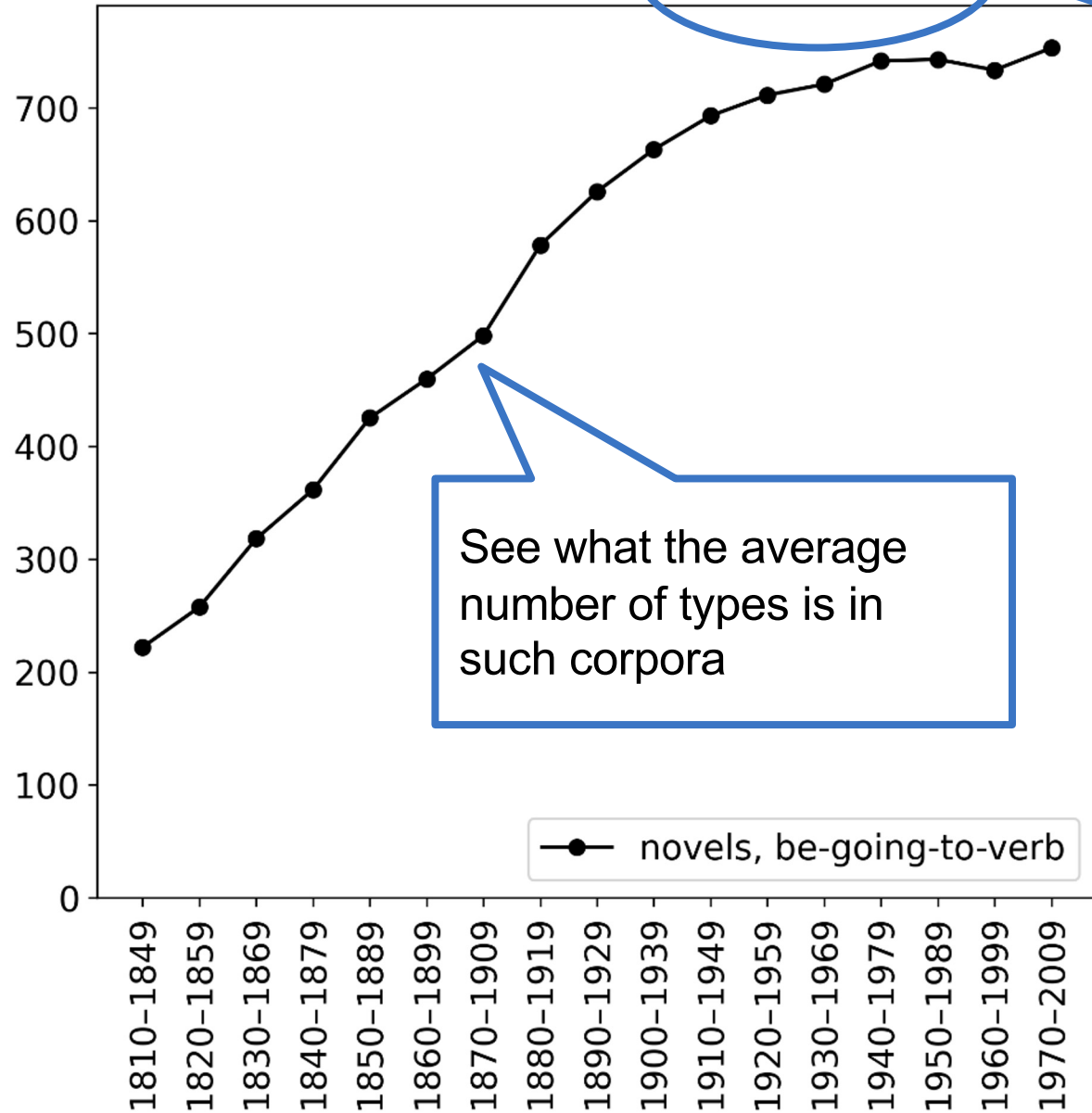
Visualizing trends

Assessing statistical significance
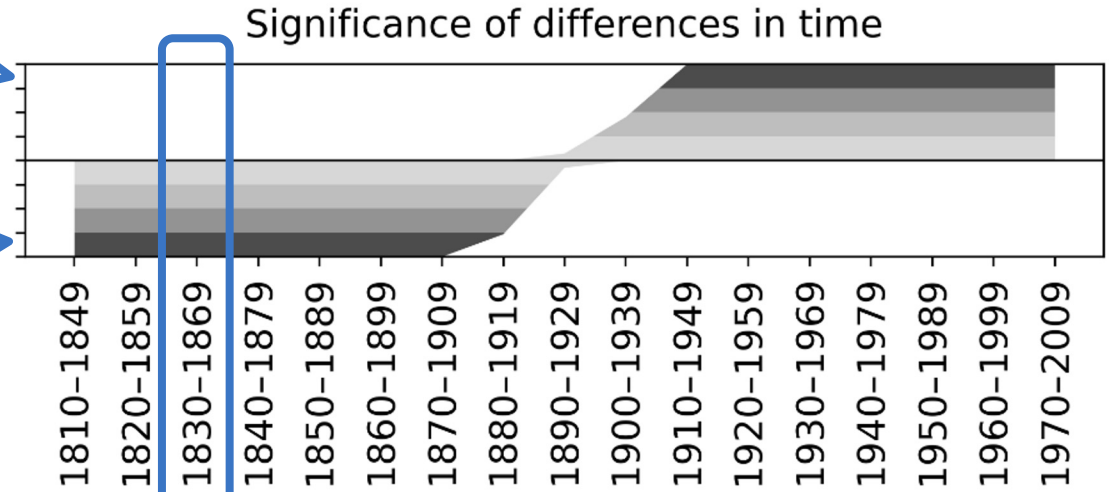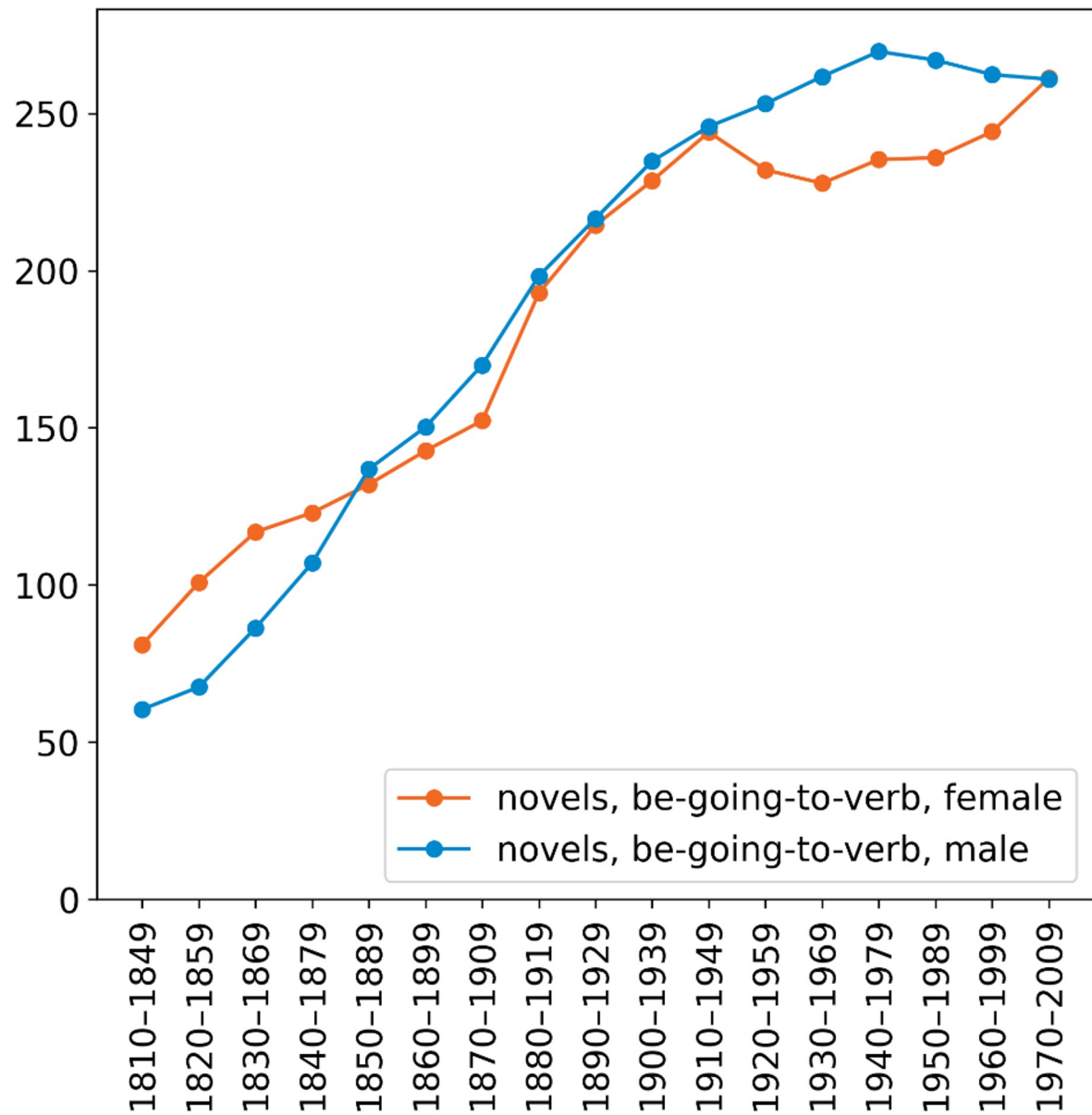
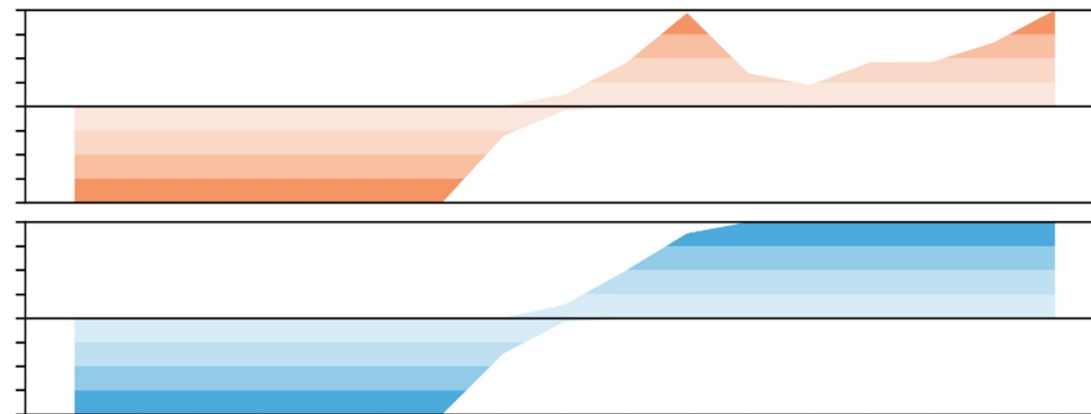Types in subcorpora with 18811353 words
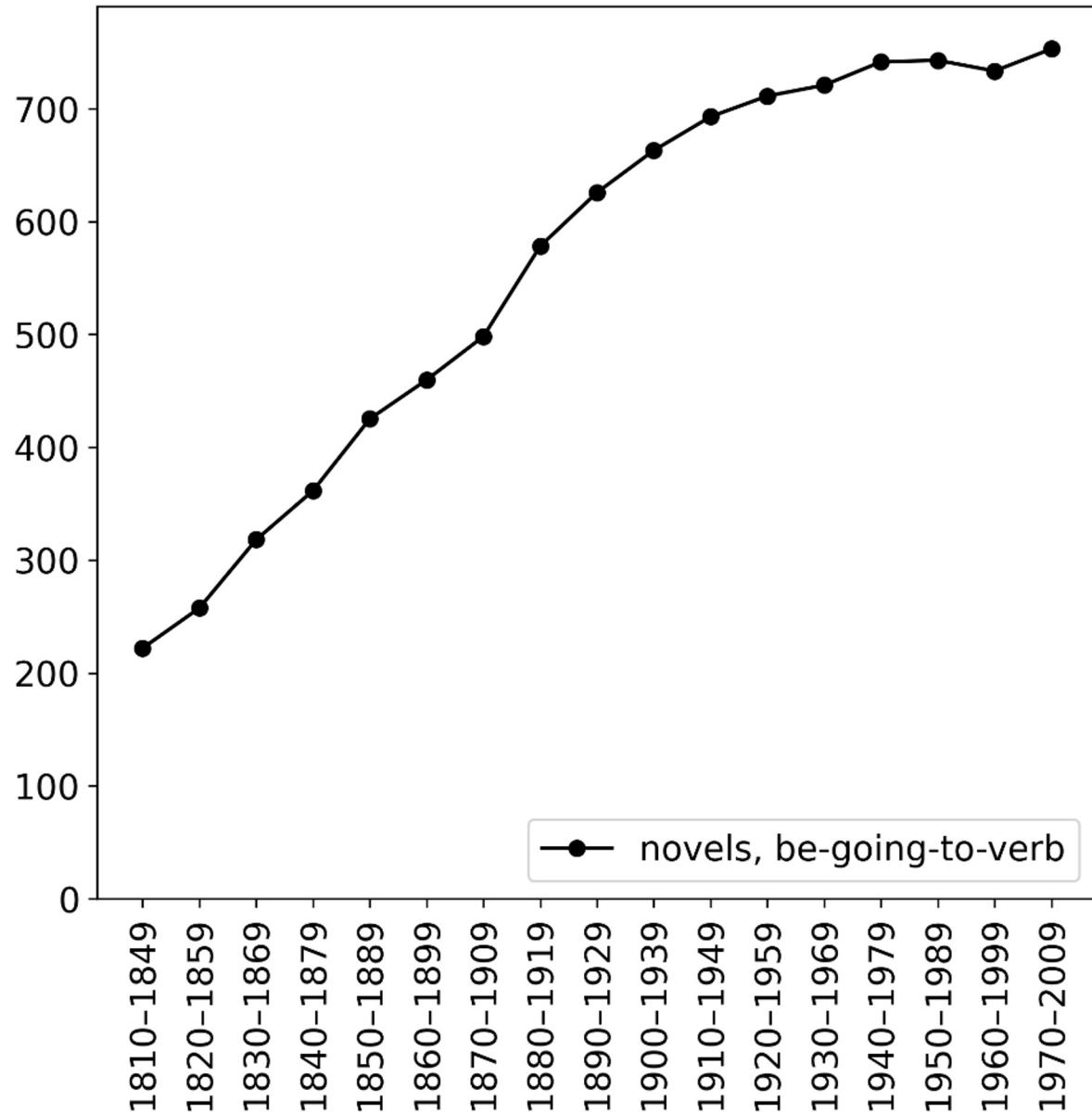
Significance of differences in time
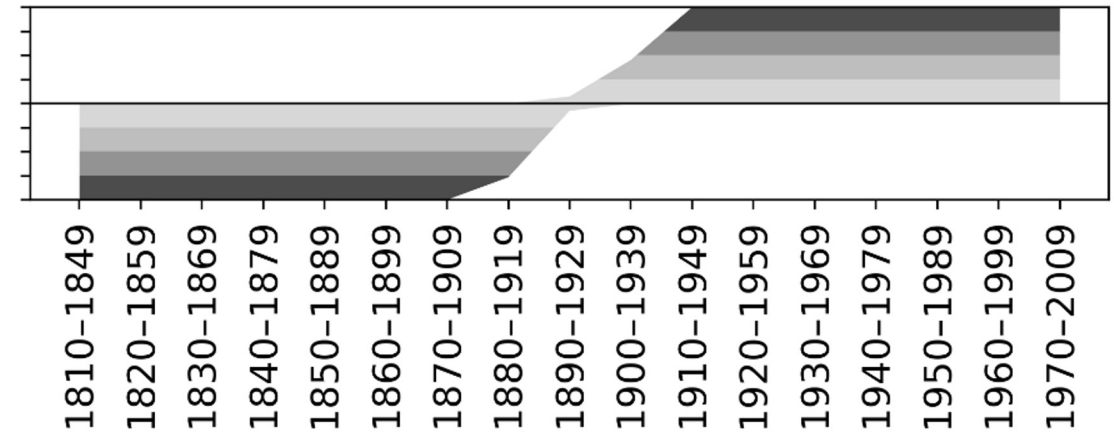
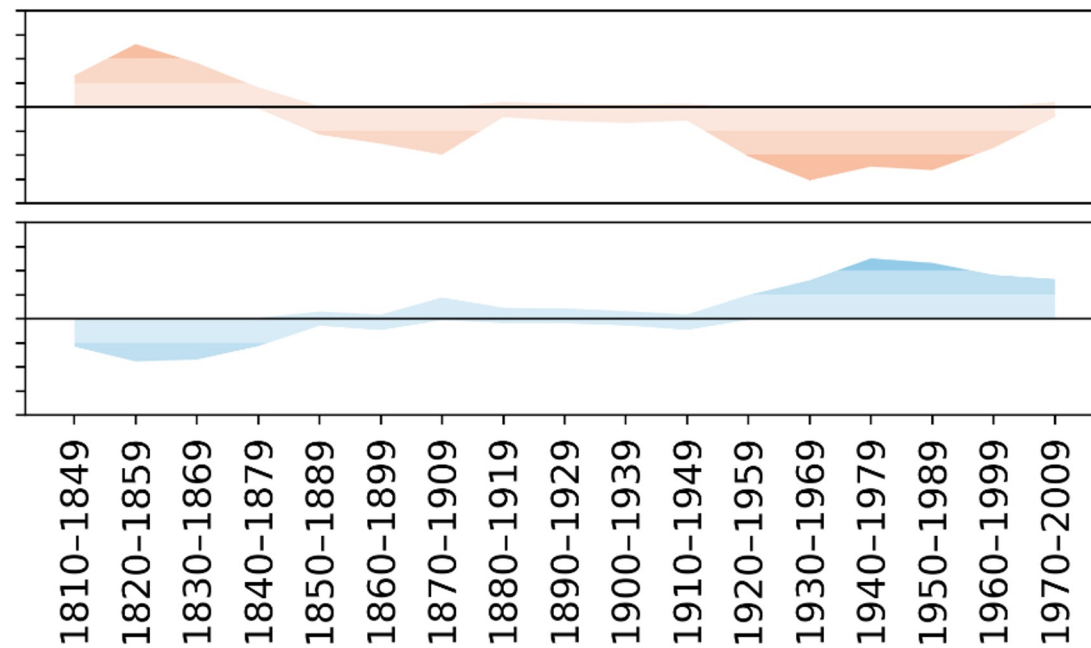A clear increasing trend that is also statistically significant
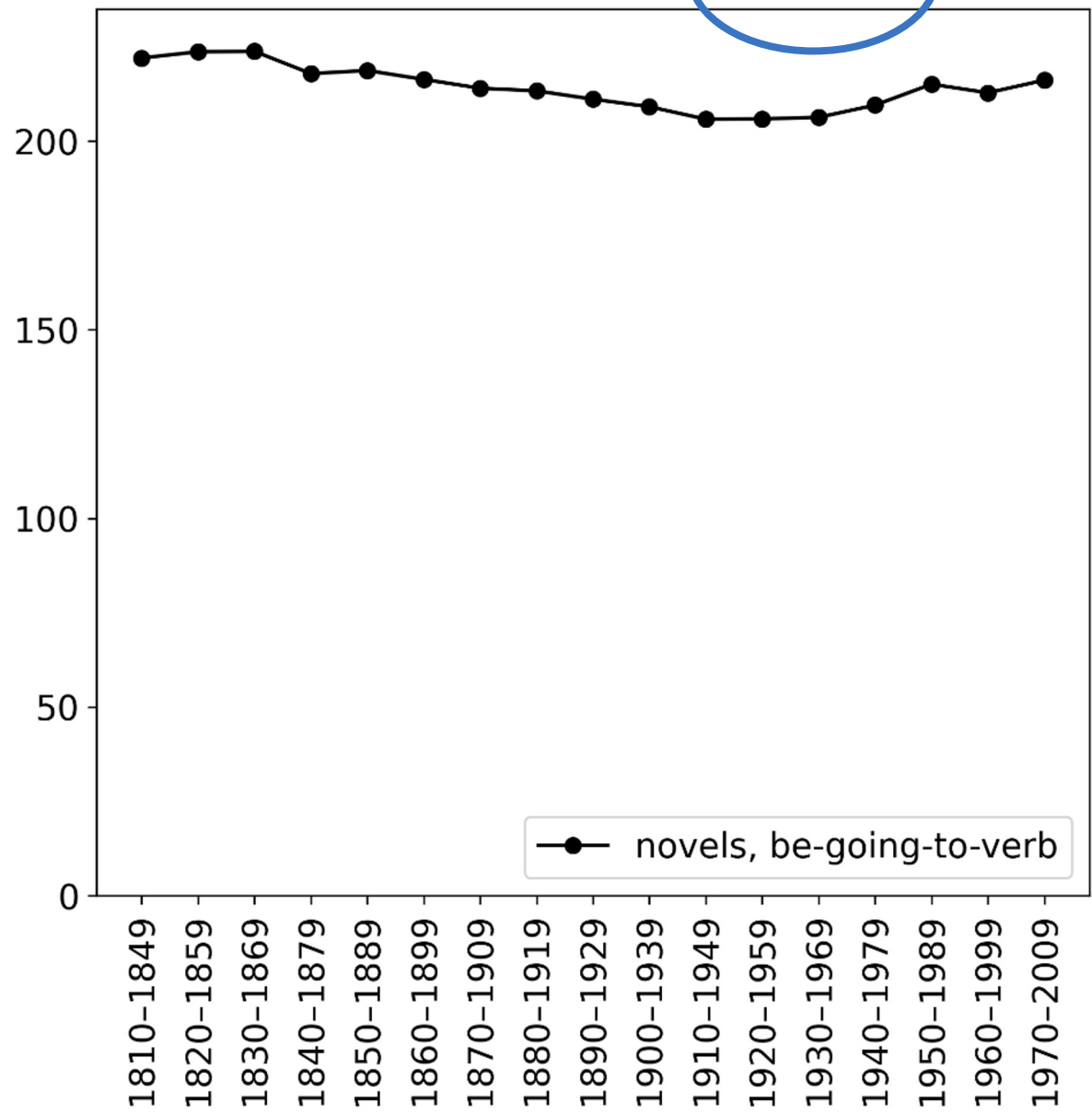
Types in subcorpora with 2863385 words

Significance of differences in time

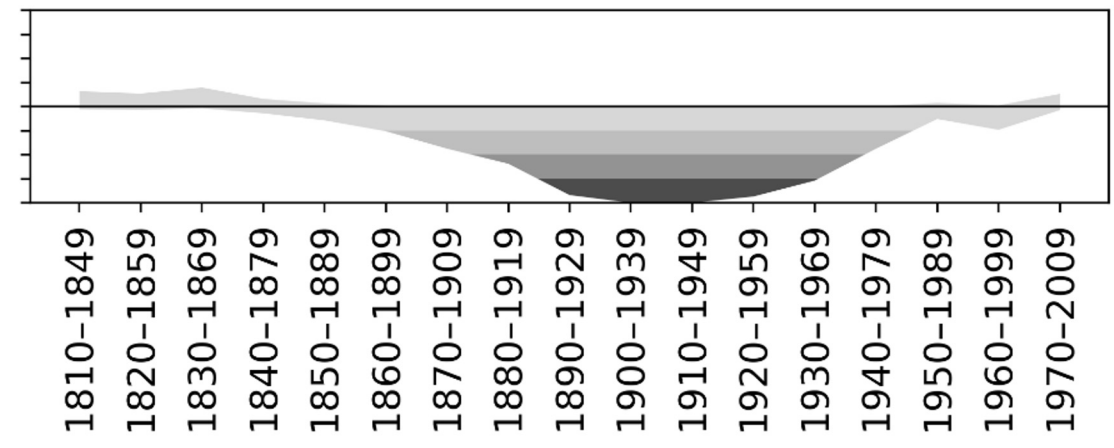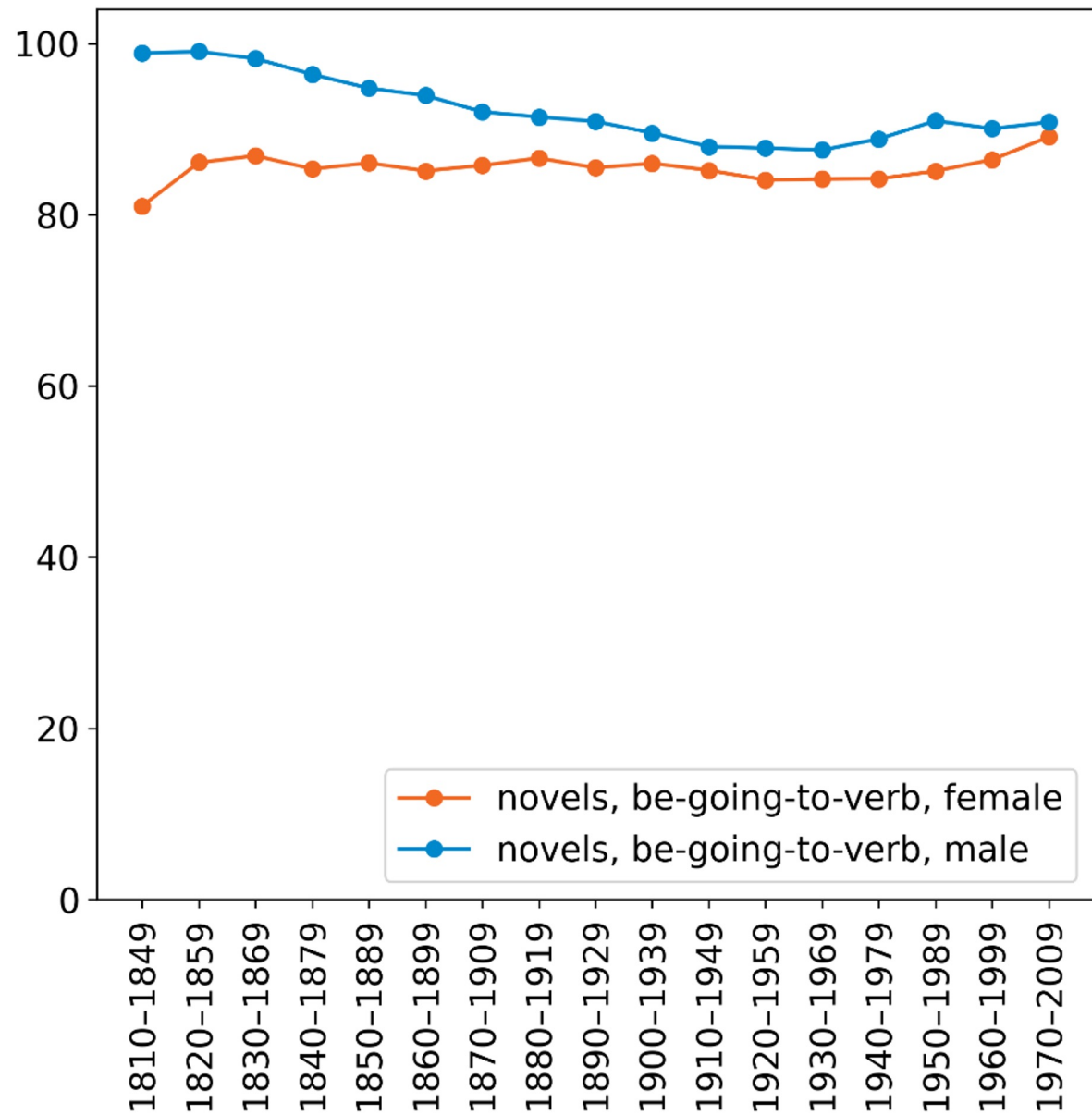Significance in comparison with other categories

Legend:
- novels, be-going-to-verb, female
- novels, be-going-to-verb, male

Types in subcorpora with 799 tokens

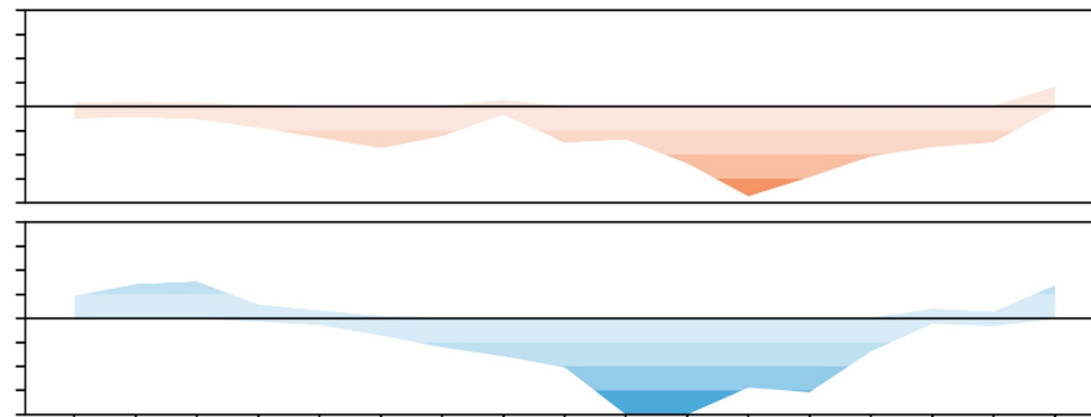Significance of differences in time
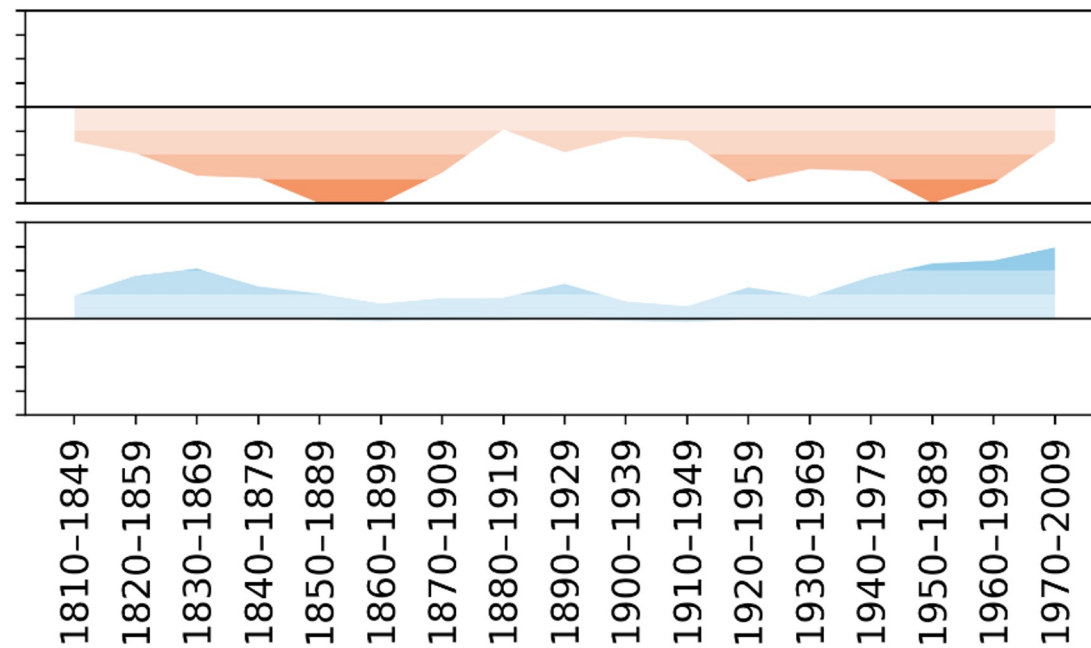
novels, be-going-to-verb

More frequent use or more diverse use?

Types in subcorpora with 212 tokens

Significance of differences in time

Significance in comparison with other categories

novels, be-going-to-verb, female
novels, be-going-to-verb, male

PROPORTION OF MENTAL VERBS
BY GENDER
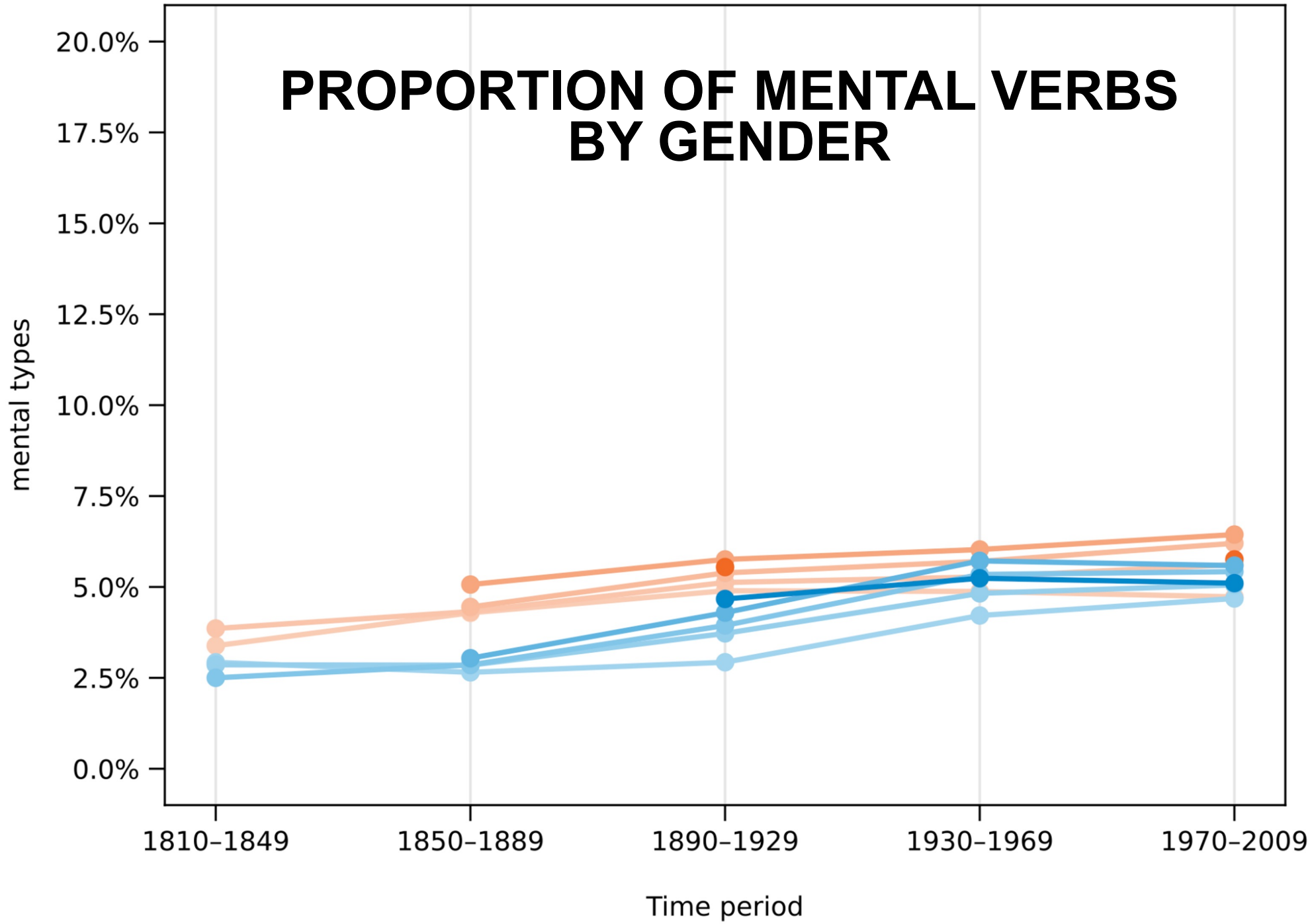
EXAMPLE

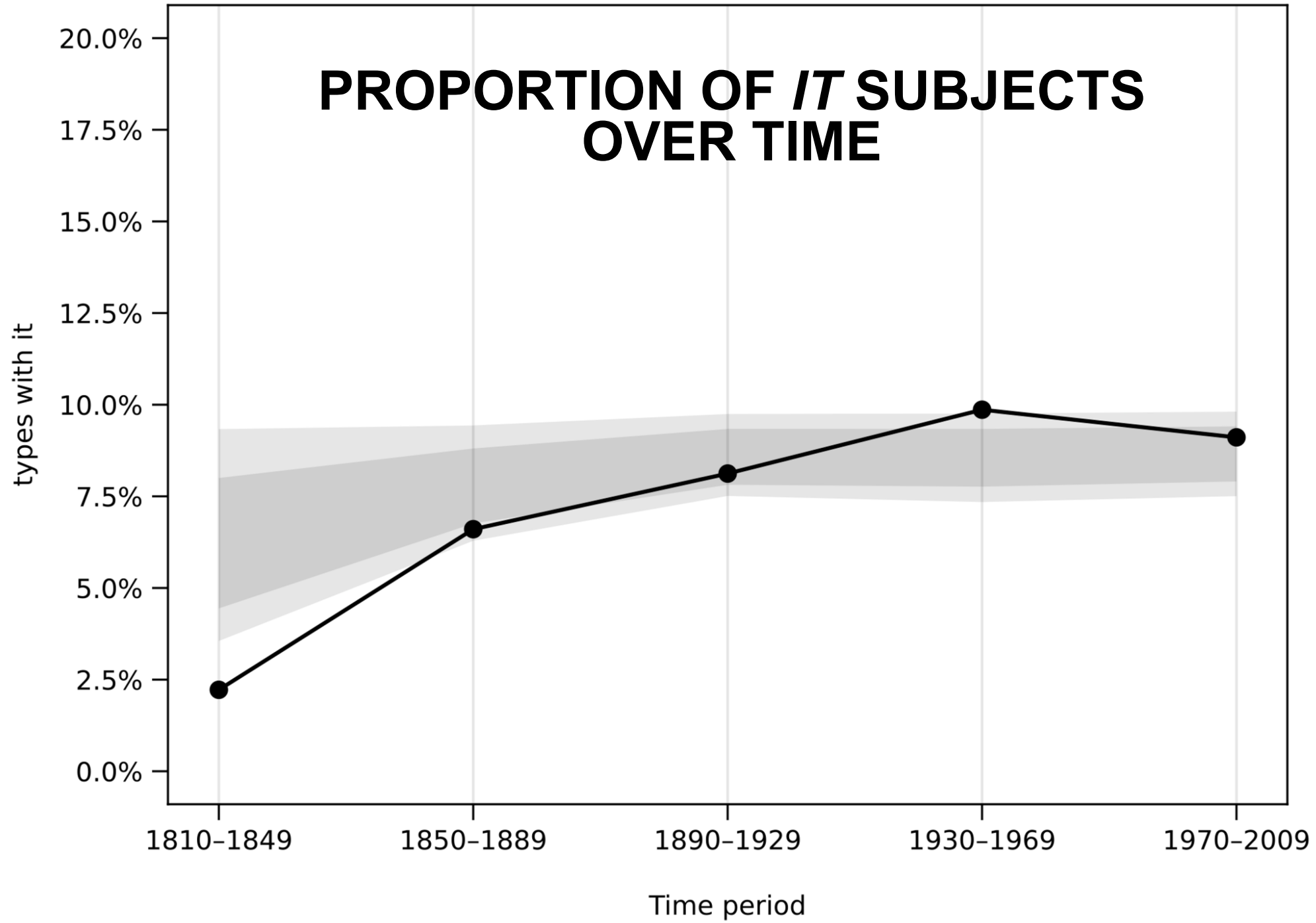Brody shook his head. "Old Hooper's going to **wish** he ate at the Abelard."

"You're a beast," she said. "Wait till you taste it. You'll change your tune."
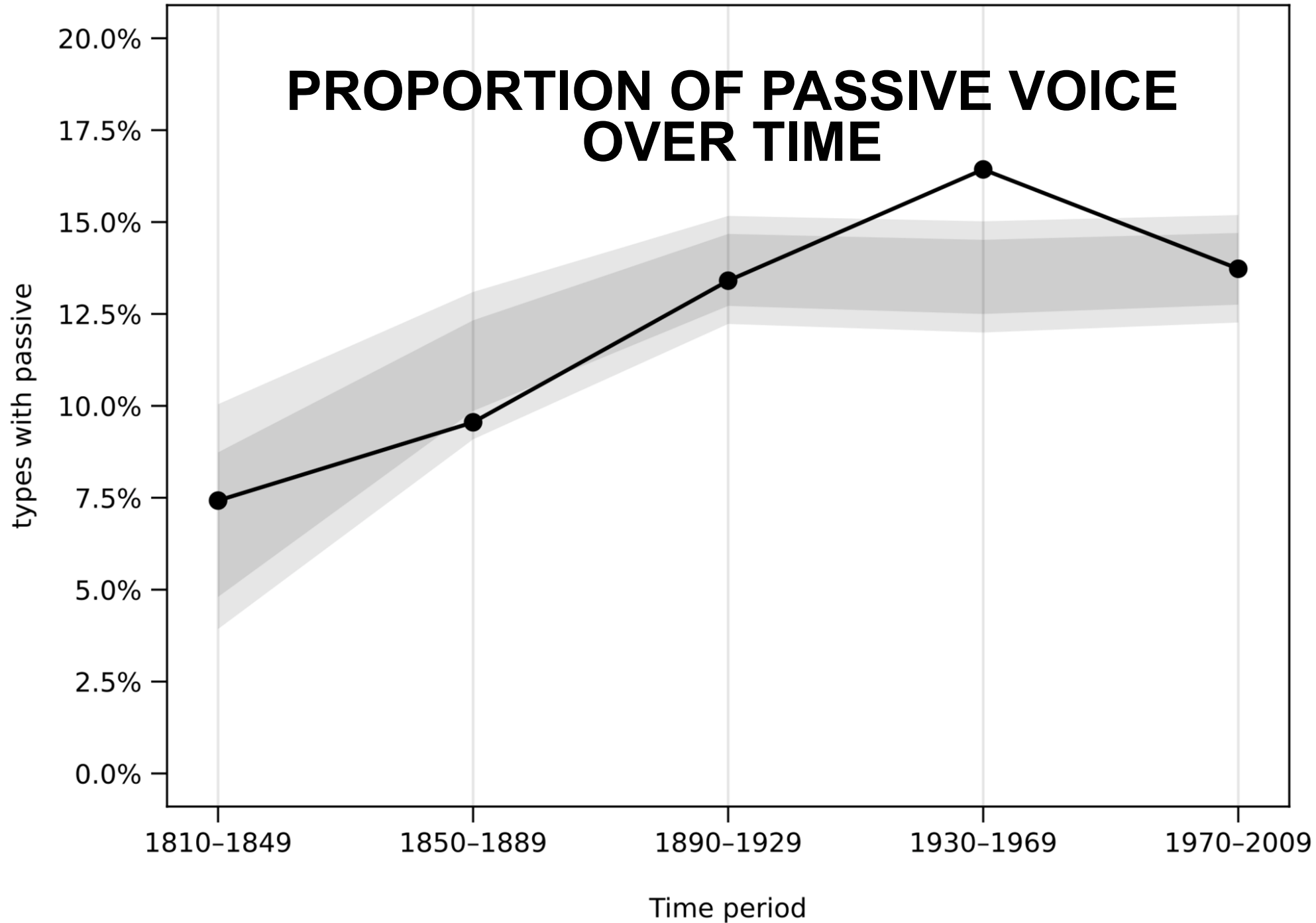
(Lillian Hellman, *Pentimento*, 1973)

Lynn Gilbert, CC BY-SA 4.0, via Wikimedia Commons

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

PROPORTION OF *IT* SUBJECTS OVER TIME

PROPORTION OF PASSIVE VOICE OVER TIME

SUMMARY OF RESULTS

- Overall productivity/type diversity of BE *going to* V doesn't increase in C19–20 AmE, even a slight decrease

  - Men's usage more productive, convergence over time

- Internal factors do indicate increasing productivity

  - Proportion of types with mental verbs (led by women), *it* subjects, passive voice

- Florent's distributional semantic analysis (not shown) identifies areas of growth

  - E.g. mental verbs, motion verbs

  - Points to an increase in grammaticalization

  - Gender differences as well, with women leading the way

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

VARIENG, Faculty of Arts

Sociolinguistic variation and change in productivity / Tanja Säily

2023-12-18

64

CONCLUSIONS

- At this stage of grammaticalization, overall type diversity stagnates but **internal factors** linked to grammaticalization indicate increasing productivity

  - Important to take into account

- Consistent **gender** differences – different leaders of change and/or different genres?

  - Gender cannot be ignored as a possible factor

  - Mental verbs could be linked to women's involved writing style (Biber & Burges 2000)

- Future work: analyse hapax legomena / new types

CONCLUSION

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

SUMMARY OF RESULTS

3 different case studies, 3 different outcomes

- *-ity*: **men lead the change**

  - Goes against findings from present-day sociolinguistics that women tend to lead change (Labov 2001: 292–293)

  - More access to formal written registers? Better match to men's informational writing style?

- very *-ed*: **no significant gender differences**

- BE *going to* V: men use a more diverse set of verbs but **women lead the change** in verb semantics (mental/motion)

  - Writing in different genres? Better match to women's involved writing style?
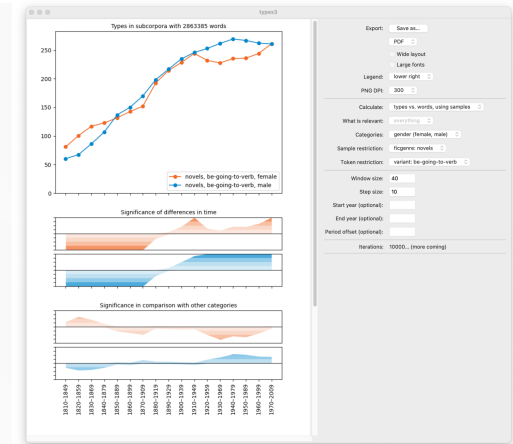
CHALLENGES WITH MATERIALS
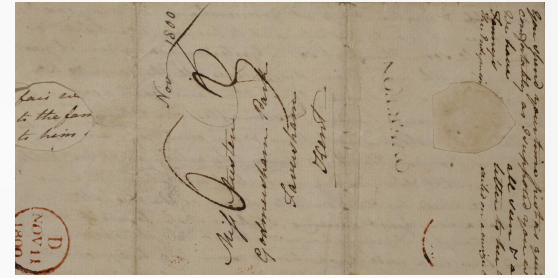
- Historical corpora that represent a wide section of society are typically **small**

  - Manuscript-based: common people (and most women) never wrote anything for publication, but anyone who was literate could write letters

  - Digitizing and annotating manuscripts is highly resource-intensive

  - → **Research limited** to:

    - Frequent phenomena, type frequencies (extent of use), few social categories (gender…)

- Large corpora typically only **represent a narrow section of society**

  - Impossible to study social class; often no metadata on other social groups, either

  - Fiction may permit analyses of gender variation, but a heterogeneous genre

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

VARIENG, Faculty of Arts

Sociolinguistic variation and change in productivity / Tanja Säily

2023-12-18

68

CHALLENGES WITH METHODS

- **No reliable multivariate method** for type-based measures

  - Cannot assess combined/relative influence of different social and internal factors

- Influence of general **vocabulary size** on results?

  - If men had more access to education, their mental lexicon would have been bigger → used a greater variety of types because of that

  - Solution 1: relate type-based measures to overall lexical diversity among social groups

  - Solution 2: focus on semantic fields rather than words

- Interactive visualization (*types*3): **no access to the texts** and metadata

  - For generating hypotheses, interpreting results

CONCLUSIONS

- There is **sociolinguistic variation and change in productivity** at different levels of language

  - Important, **understudied area**; sociolinguistics has focused on token frequencies

  - Principles of linguistic change (e.g. Labov 2001) may need to be revised based on this research

- **Better materials and methods** are being developed

  - New AI methods may help with transcribing + annotating manuscript-based corpora

  - Multidisciplinary collaboration

REFERENCES

- Baayen, R.H. 1992. Quantitative aspects of morphological productivity. Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1991*, 109–149. Kluwer.

- Baayen, R.H. 1993. On frequency, transparency and productivity. Geert Booij & Jaap Van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Kluwer.

- Baayen, R.H. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 899–919. Mouton de Gruyter.

- Biber, Douglas & Jená Burges. 2000. Historical change in the language use of women and men: Gender differences in dramatic dialogue. *Journal of English Linguistics* 28(1): 21–37.

- Bolinger, Dwight L. 1948. On defining the morpheme. *Word 4:* 18–23.

- Brezina, Vaclav. 2005. *The development of the prefixes un- and in- in Early Modern English with special regard to the sociolinguistic background*. Faculty of Arts, Charles University in Prague MA thesis.

- Budts, Sara & Peter Petré. 2016. Reading the intentions of *be going to*: On the subjectification of future markers. *Folia Linguistica Historica* 37: 1–32.

- Cowie, Claire & Christiane Dalton-Puffer. 2002. Diachronic word-formation and studying changes in productivity over time: Theoretical and methodological considerations. Javier E. Díaz Vera (ed.), *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics*, 410–437. Rodopi.

- Denison, David. 1998. Syntax. In Suzanne Romaine (ed.), *The Cambridge history of the English language, vol. 4, 1776–1997*, 92–329. CUP.

- Halliday, M.A.K. & Christian M.I.M. Matthiessen (2014). *Halliday's introduction to functional grammar*, 4th edition. Routledge.

- Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. CUP.

- Hilpert, Martin & Florent Perek. 2022. You don't get to see that every day: On the development of permissive *get*. *Constructions and Frames* 14(1): 14–41.

- Keune, Karen. 2012. *Explaining register and sociolinguistic variation in the lexicon: Corpus studies on Dutch*. Radboud University Nijmegen PhD dissertation.

- Keune, Karen, Roeland van Hout & R. H. Baayen. 2006. Socio-geographic variation in morphological productivity in spoken Dutch: a comparison of statistical techniques. Jean-Marie Viprey & Lexicometrica (eds.), *Actes de JADT 2006 : 8es journées internationales d'analyse statistique des données textuelles*, 571–581. Université de Franche-Comté.

- Labov, William. 2001. *Principles of linguistic change, volume 2: Social factors*. Blackwell.

- Lupica Spagnolo, Marta. 2013. Morphologische Produktivität in deutschsprachigen Texten nicht nativer Autoren: eine korpuslinguistische Analyse. *Zeitschrift für germanistische Linguistik* 41(3): 339–376.

- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. Pearson Education.

- Öhman, Emily, Tanja Säily & Mikko Laitinen. 2019. Towards the inevitable demise of everybody? A multifactorial analysis of *-one/-body/-man* variation in indefinite pronouns in historical American English. Paper presented at ICAME 40, Neuchâtel, Switzerland, June 2019. https://tanjasaily.fi/talks/icame40_ohman_et_al_2019.pdf

REFERENCES

- Palmer, Chris C. 2009. *Borrowings, derivational morphology, and perceived productivity in English, 1300–1600*. University of Michigan PhD dissertation.

- Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1): 149–188.

- Perek, Florent. 2018. Recent change in the productivity and schematicity of the *way*-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1): 65-97.

- Rodríguez-Puente, Paula. 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, Paula, Tanja Säily & Jukka Suomela. 2022. New methods for analysing diachronic suffix competition across registers: How *-ity* gained ground on -*ness* in Early Modern English. *International Journal of Corpus Linguistics* 27(4): 506–528.

- Romaine, Suzanne. 1983. On the productivity of word formation rules and limits of variability in the lexicon. *Australian Journal of Linguistics* 3(2): 177–200.

- Säily, Tanja. 2014. *Sociolinguistic variation in English derivational productivity: Studies and methods in diachronic corpus linguistics*. Société Néophilologique.

- Säily, Tanja, Victorina González-Díaz & Jukka Suomela. 2018. Variation in the productivity of adjective comparison in Present-day English. In Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014*, 159–184. Routledge.

- Säily, Tanja, Martin Hilpert & Jukka Suomela. Forthcoming. New approaches to investigating change in derivational productivity: Gender and internal factors in the development of *-ity* and -*ness*, 1600–1800. In Patricia Ronan, Theresa Neumaier, Lisa Westermayer, Andreas Weilinghoff & Sarah Buschfeld (eds.), *Crossing boundaries through corpora: Innovative approaches to corpus linguistics*. Benjamins.

- Säily, Tanja & Jukka Suomela. 2009. Comparing type counts: The case of women, men and *-ity* in early English letters. In Antoinette Renouf & Andrew Kehoe (eds.), *Corpus linguistics: Refinements and reassessments*, 87–109. Rodopi.

- Säily, Tanja & Jukka Suomela. 2017. *types*2: exploring word-frequency differences in corpora. In Turo Hiltunen, Joe McVeigh & Tanja Säily (eds.), *Big and rich data in English corpus linguistics: Methods and explorations*. VARIENG. https://urn.fi/URN:NBN:fi:varieng:series-19-10

- Säily, Tanja & Turo Vartiainen. Forthcoming. Historical linguistics. In Michaela Mahlberg & Gavin Brooks (eds.), *Bloomsbury handbook of corpus linguistics*. Bloomsbury.

- Säily, Tanja, Turo Vartiainen & Harri Siirtola. 2017. Exploring part-of-speech frequencies in a sociohistorical corpus of English. Tanja Säily, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds.), *Exploring Future Paths for Historical Sociolinguistics*, 23–52. Benjamins.

- Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6): 1237–1269.

- Vartiainen, Turo. 2016. A constructionist approach to category change: Constraining factors in the adjectivization of participles. *Journal of English Linguistics* 44(1): 34–60.

- Vartiainen, Turo. 2021. Trends and recent change in the syntactic distribution of degree modifiers: Implications for a usage-based theory of word classes. *Journal of English Linguistics* 49(2): 228–251.

- Wu, Junhui, Qingshun He & Guangwu Feng. 2016. Rethinking the grammaticalization of future *be going to*: A corpus-based approach. *Journal of Quantitative Linguistics* 23(4): 317–341.

BONUS SLIDES

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

LANGUAGE VARIATION AND CHANGE / SOCIOLINGUISTICS

- Labov (2001) analysed the **influence of social factors on linguistic change**

  - Basis: studies of sound change in contemporary American English

  - Used **apparent-time data**, i.e. compared speakers of different ages, assuming that speakers represent the language use of the time they were born

- **Gender paradox**: "Women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not" (Labov 2001: 293)

  - Women lead **changes from above** the level of social awareness, when the linguistic feature in question is overtly prescribed

  - But they also lead **changes from below** the level of social awareness, when the feature is yet to attract the attention of the speech community

LANGUAGE VARIATION AND CHANGE / *HISTORICAL* SOCIOLINGUISTICS

- Uses **real-time data**, i.e. compares language use at different periods in history, often over hundreds of years → a more complete picture of (morphosyntactic) change

- Has put Labov's principles to the test

  - Gender paradox: **changes from above not always led by women**; in some cases women have had less access to the norms in question, so that men have led "changes that emanated from the world of learning and professional use" (Nevalainen & Raumolin-Brunberg 2003: 131)

- Variation and change in **productivity** not yet studied much from a sociolinguistic perspective
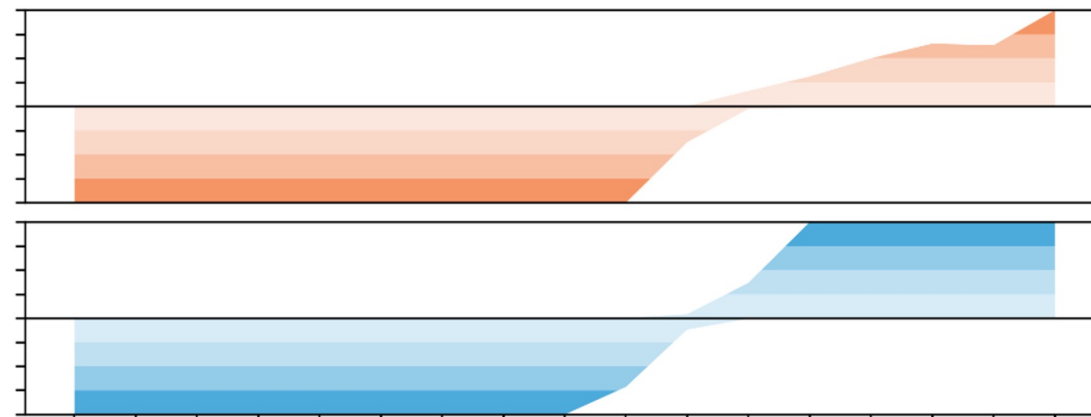
WHAT ABOUT *GONNA*?
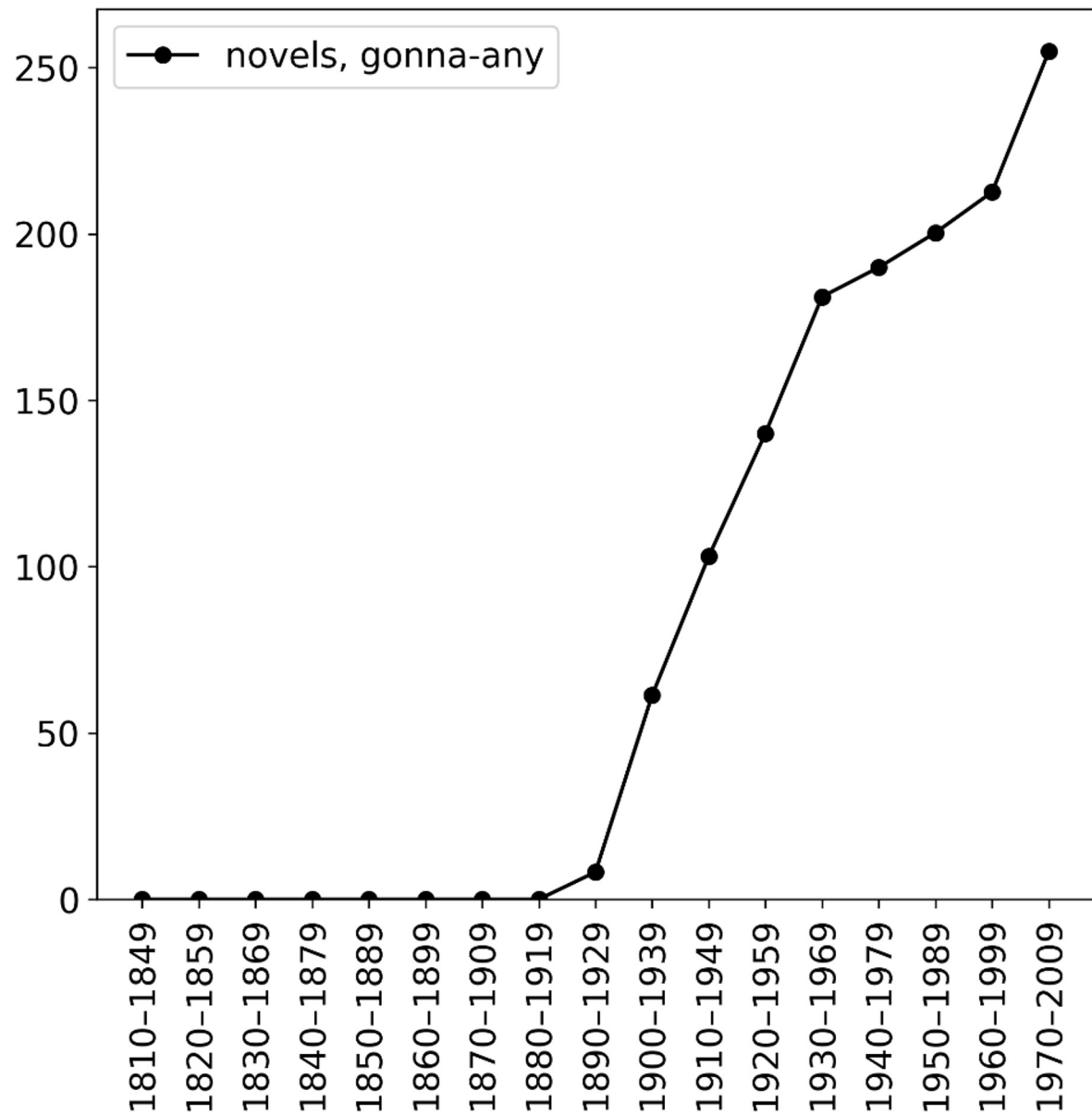
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Types in subcorpora with 18811353 words

Significance of differences in time

Types in subcorpora with 2863385 words

- novels, gonna-any, female
- novels, gonna-any, male
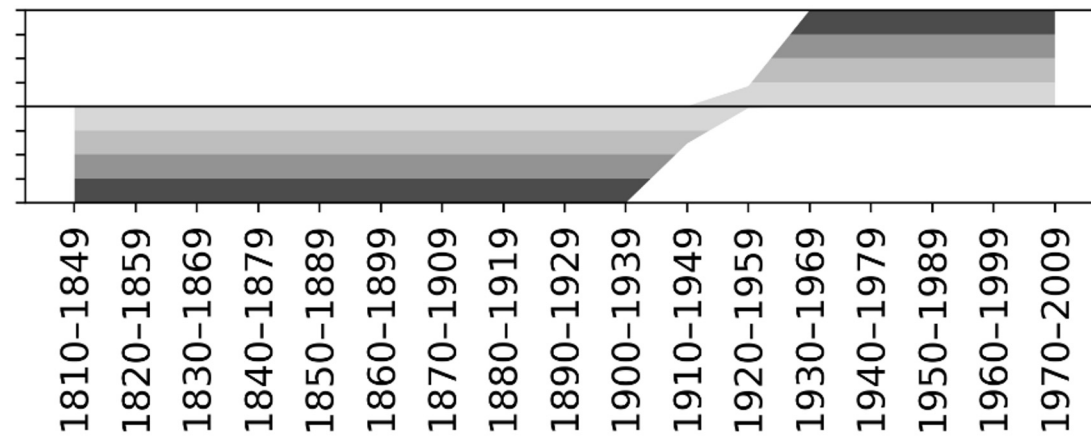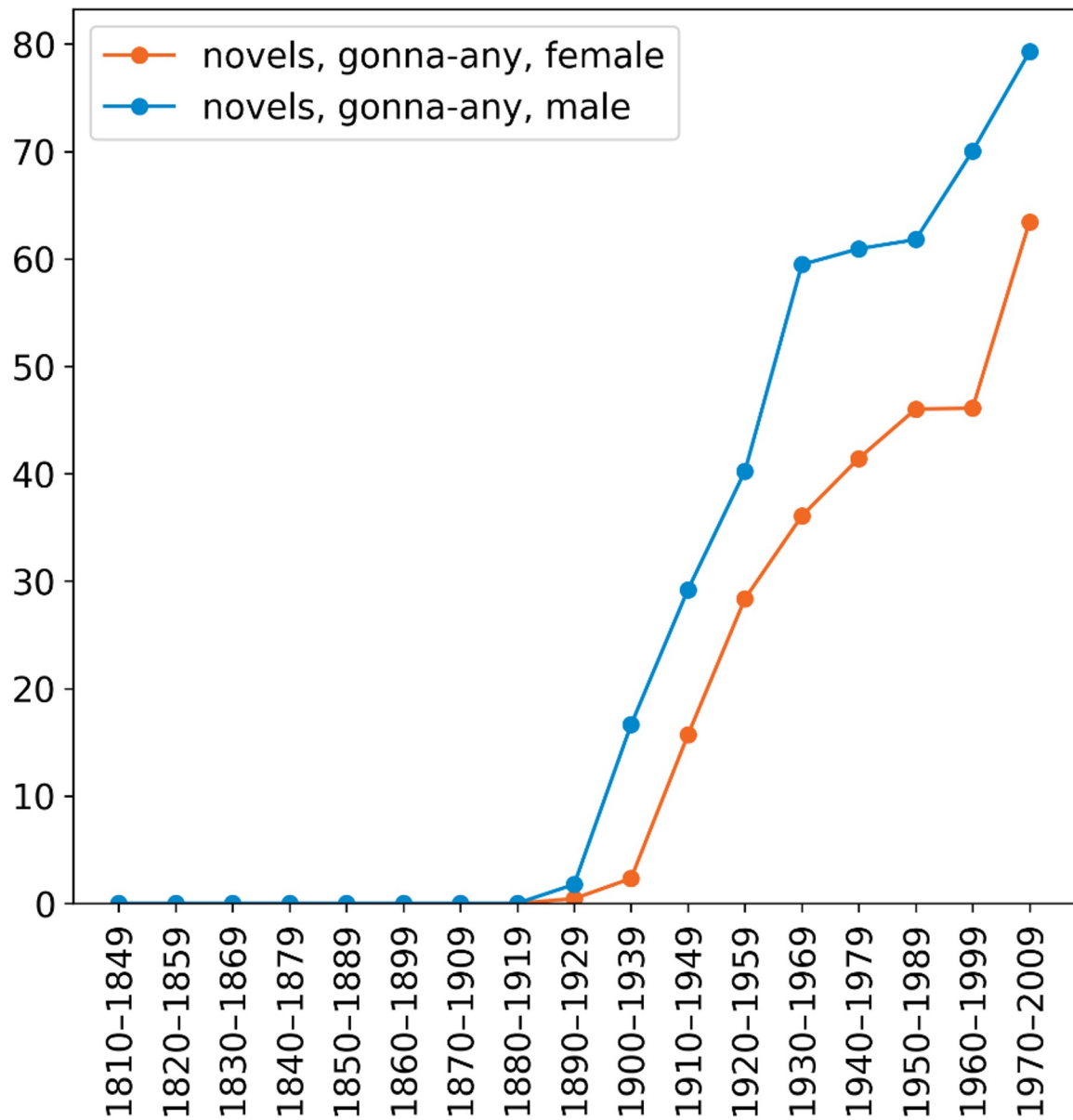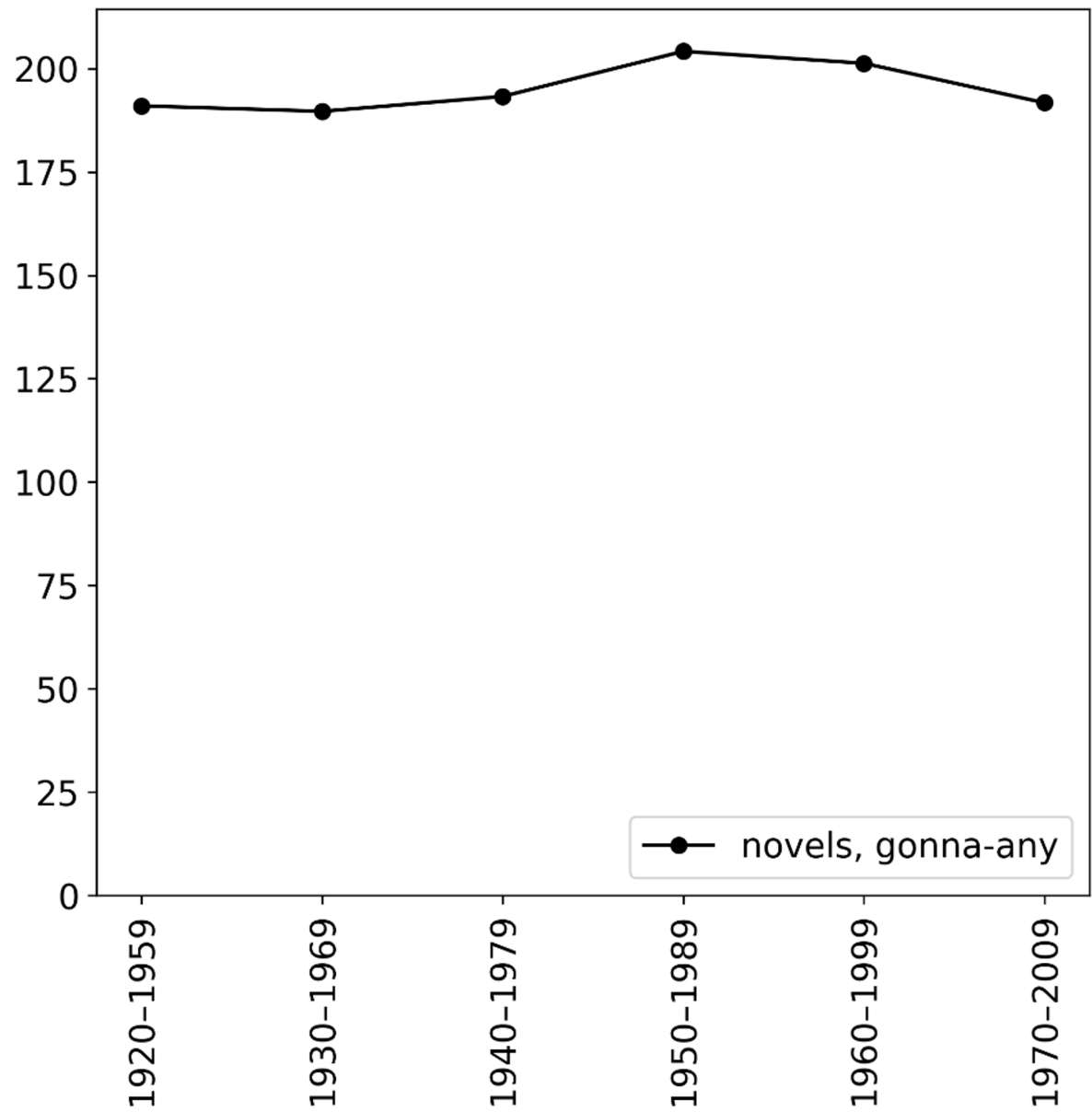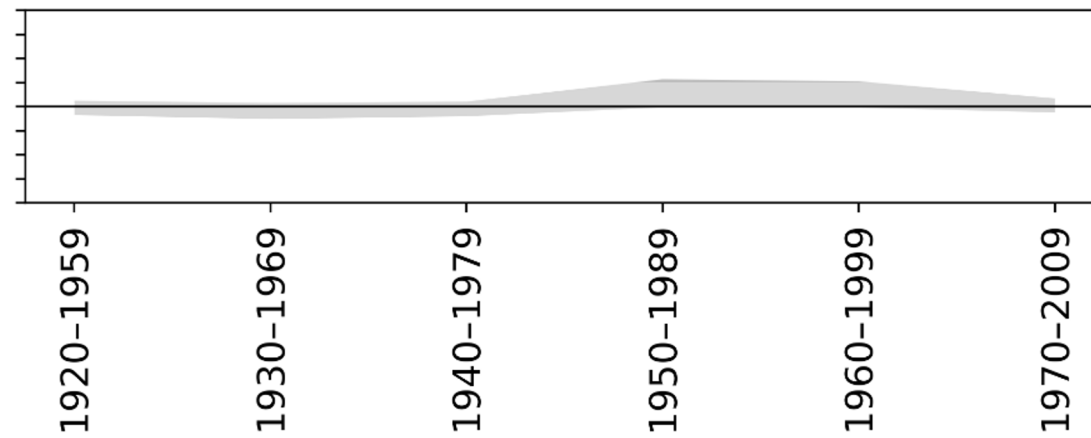
Significance of differences in time

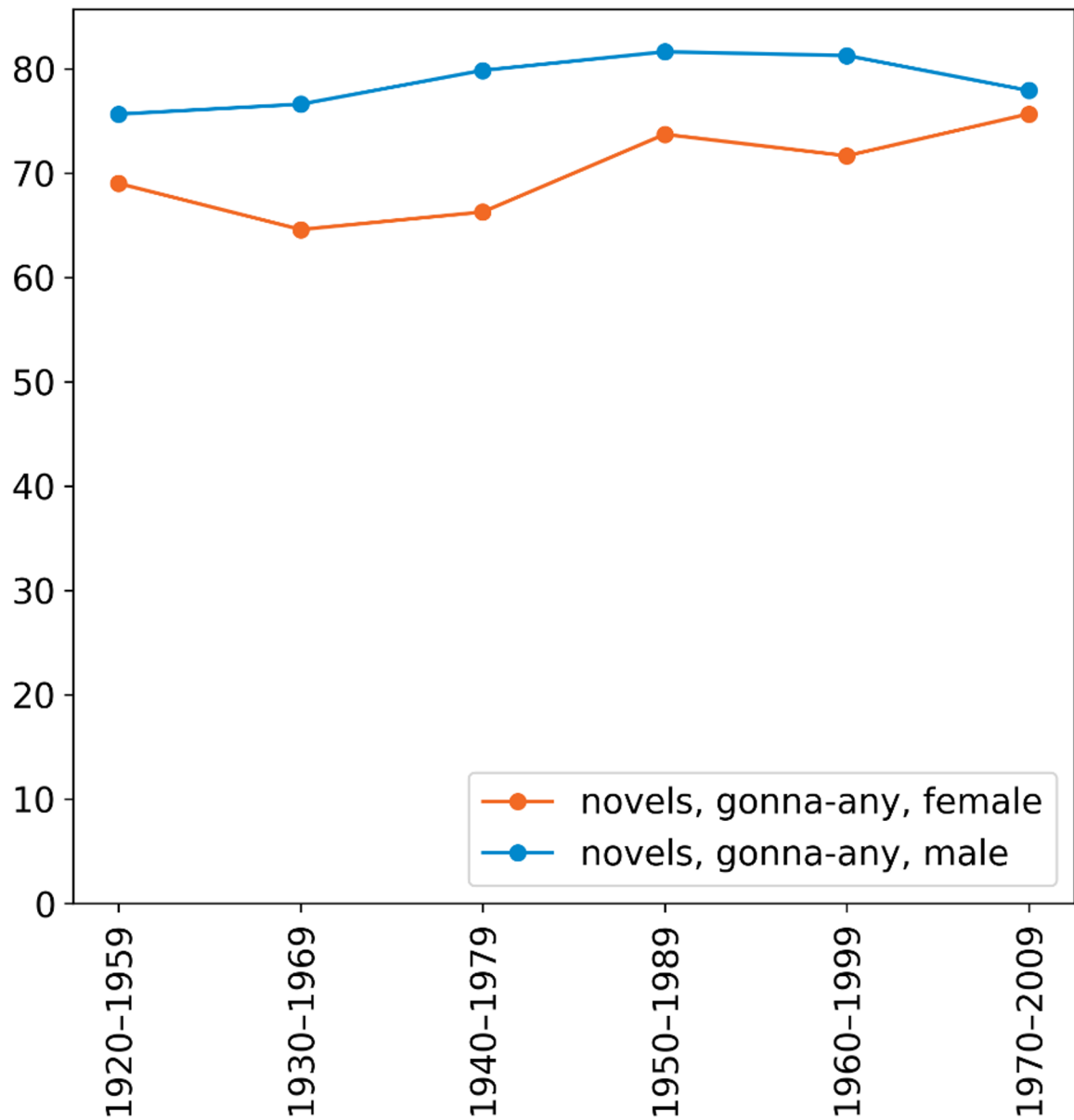**Significance in comparison with other categories**
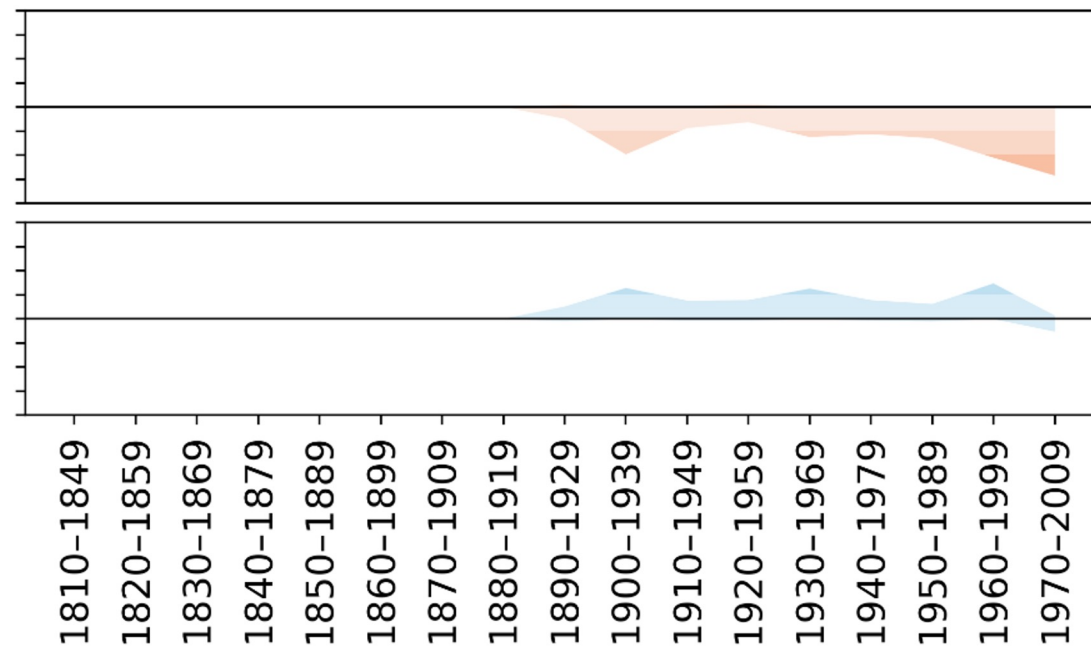
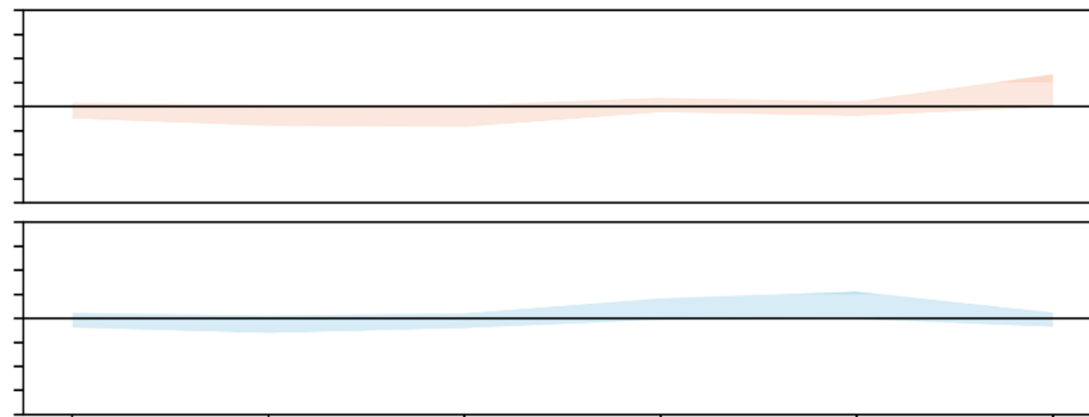Types in subcorpora with 771 tokens — Significance of differences in time

Types in subcorpora with 174 tokens

Significance of differences in time

Significance in comparison with other categories

- novels, gonna-any, female
- novels, gonna-any, male