

# Historical sociolinguistics meets constructional change

Gender and the *way*-construction in the *Corpus of Historical American English*

Florent Perek, Tanja Säily & Jukka Suomela

# CxG meets Historical Sociolinguistics

— — —

- Many previous studies suggest that language change can be driven by particular social groups, e.g. women vs. men (Nevalainen & Raumolin-Brunberg 2003)
- Here we investigate gender differences in productivity (Säily 2011)
- **Productivity** = change in the range of lexical items that can be used in a construction

# The *way*-construction

---

- Verb + Possessive + *way* + PP

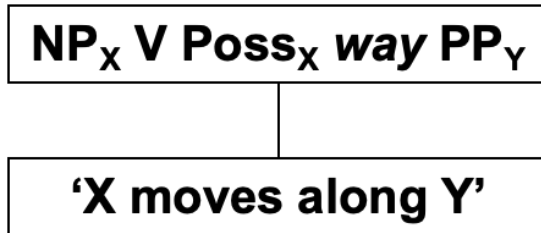
*They hacked their way through the jungle.*

*We pushed our way into the bar.*

- We focus on the “path-creation” sense: the verb refers to the means that enable motion (Traugott & Trousdale 2013, Perek 2018)
- Vs. “manner” sense or “incidental-action” sense:

*They trudged their way through the snow.*

*He whistled his way across the room.*



# The *way*-construction

— — —

- Path-creation sense claimed to originate from transitive uses of verbs with the noun *way* (Israel 1996, Traugott & Trousdale 2013)
- Convey the literal creation or maintenance of a path, e.g. with *dig*, *pave*
- Motion along that path is initially an implicature
- Constructionalization: path-creation and motion directly paired with “V one’s way PP” (Traugott & Trousdale 2013)

# The *way*-construction

---

- Construction initially centered on physical verbs, in line with the diachronic origin (Israel 1996, from OED data)  
*cut, hew, sheer, plough, dig, clear*, etc. (17th–18th century)
- Other, more abstract verbs are attested later, especially from the 19th century onwards

*smirk, spell, write* (Israel 1996)

*joke, laugh, talk, bully* (Perek 2018)

# The *way*-construction

---

- Perek (2018): increasingly wide range of verbs in 19th-20th AmE (COHA), away from concrete path-creation
- Many new verb classes correspond to unusual ways to create a path: interaction, commerce, cognition, etc.
- More likely to involve abstract, metaphorical motion (Perek 2018, 2020)

*They talk about Uncle Paul having **bought his way into** the Senate!*

# Research questions

— — —

1. Did the social factor of **gender** play a role in the changes in the productivity of the construction?
2. Are there differences in the semantic areas favoured by men vs. women at different points in time?

# 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**

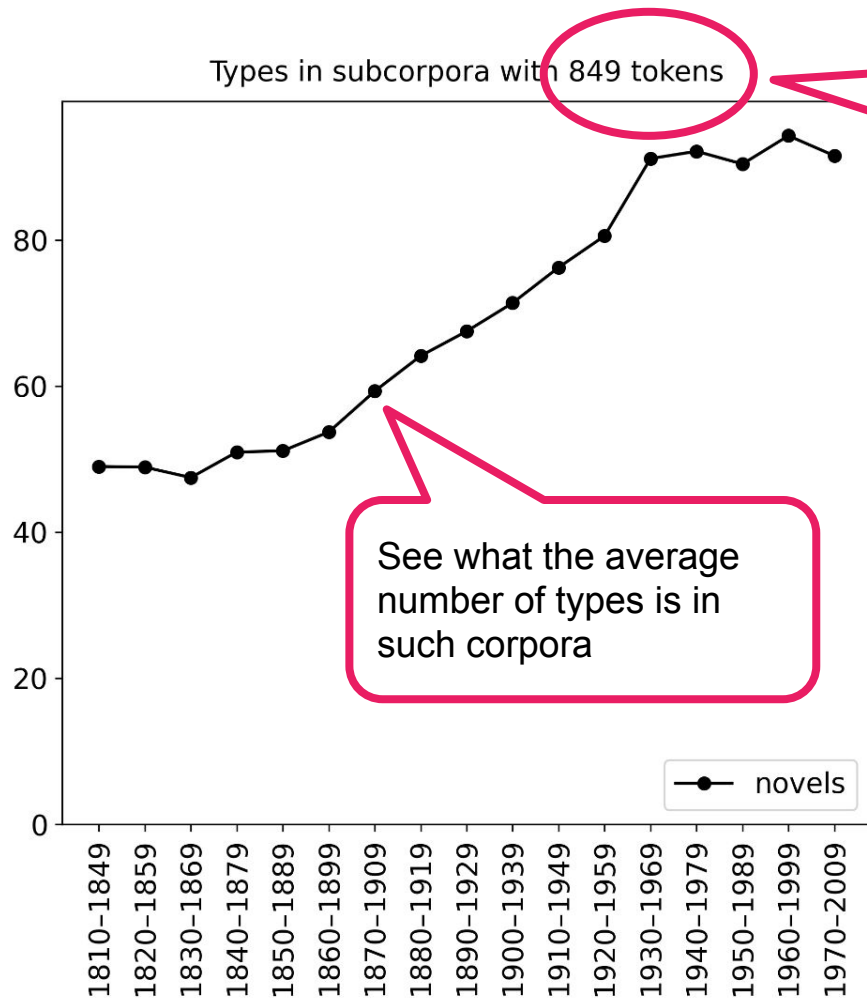


# Analysis 1: type frequencies

# Methods

— — —

- We study the productivity of the *way*-construction (path-creation sense) by measuring **type frequencies**
  - I.e. how many different items in the verb slot 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**?

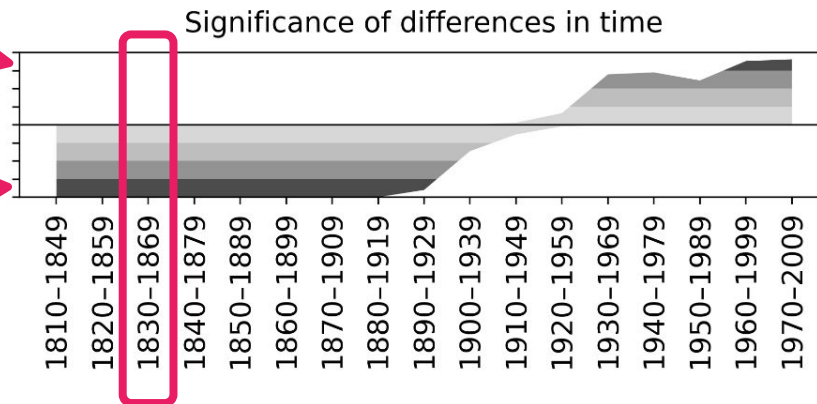


Visualizing  
trends

# Assessing statistical significance

These periods have significantly many types

These periods have significantly few types

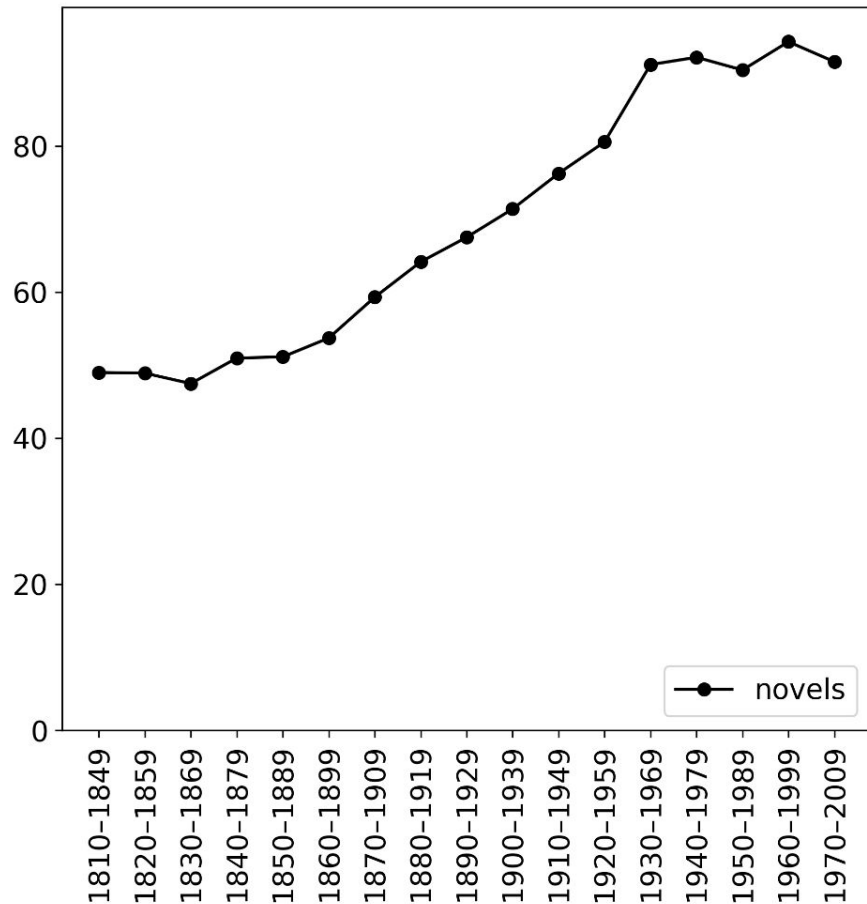


For each period (using **all** of the data):

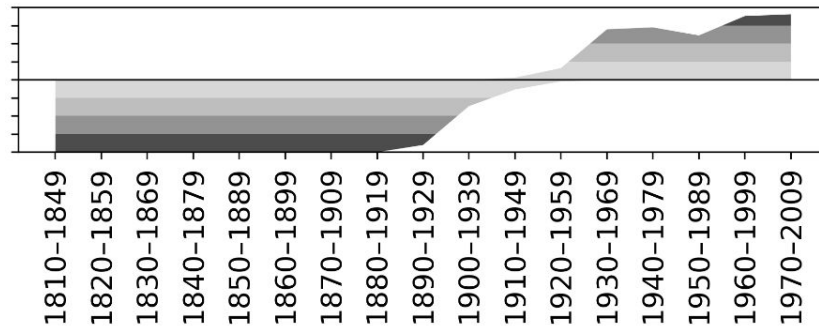
Sample random subcorpora from the whole corpus until you have a subcorpus of a comparable size

Do you typically get more or fewer types?

Types in subcorpora with 849 tokens

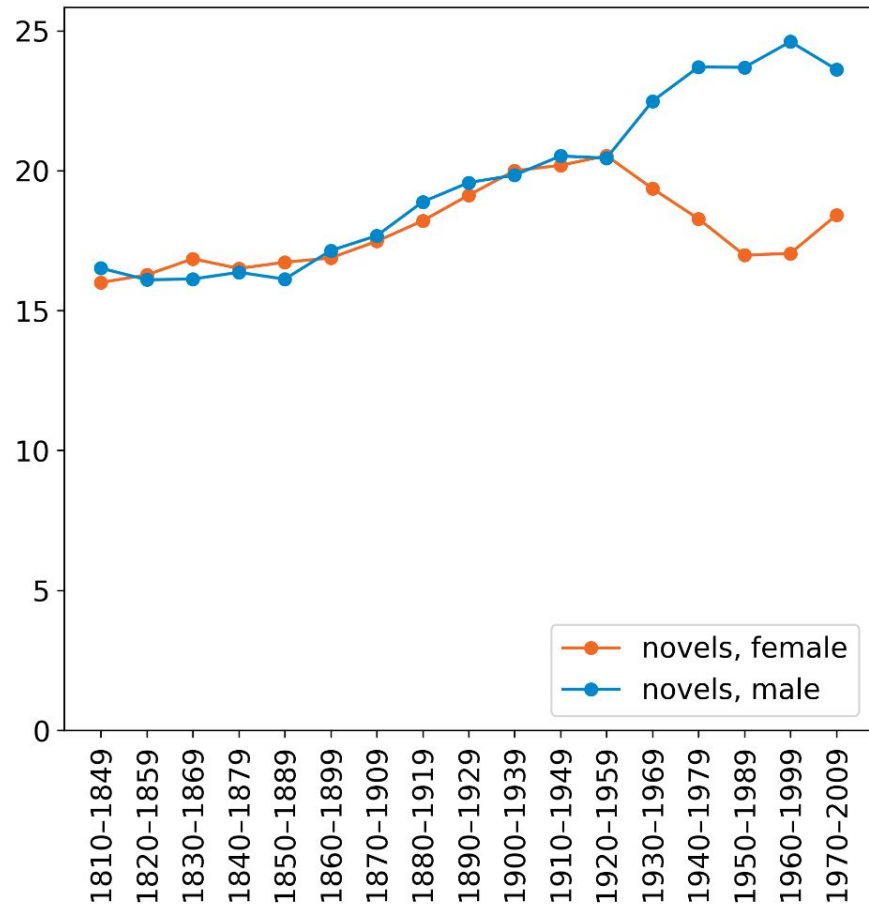


Significance of differences in time

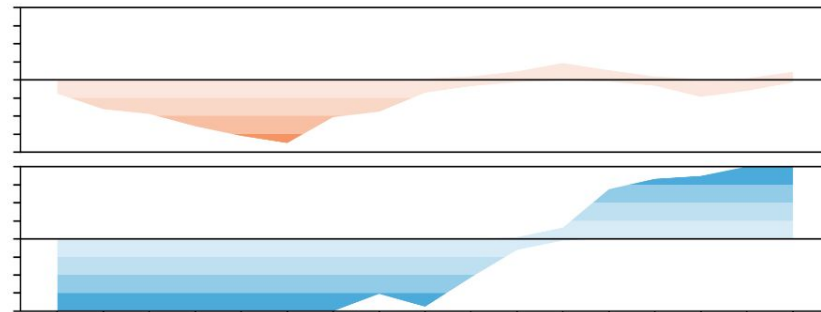


A clear increasing trend that is also statistically significant

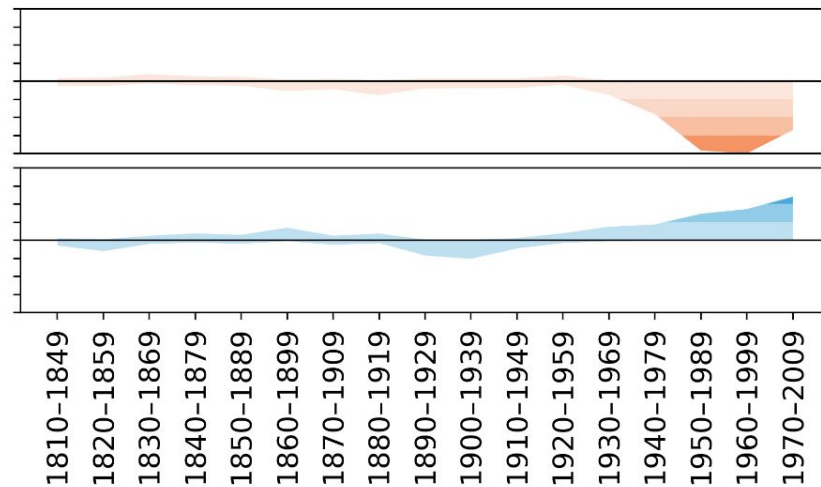
Types in subcorpora with 79 tokens



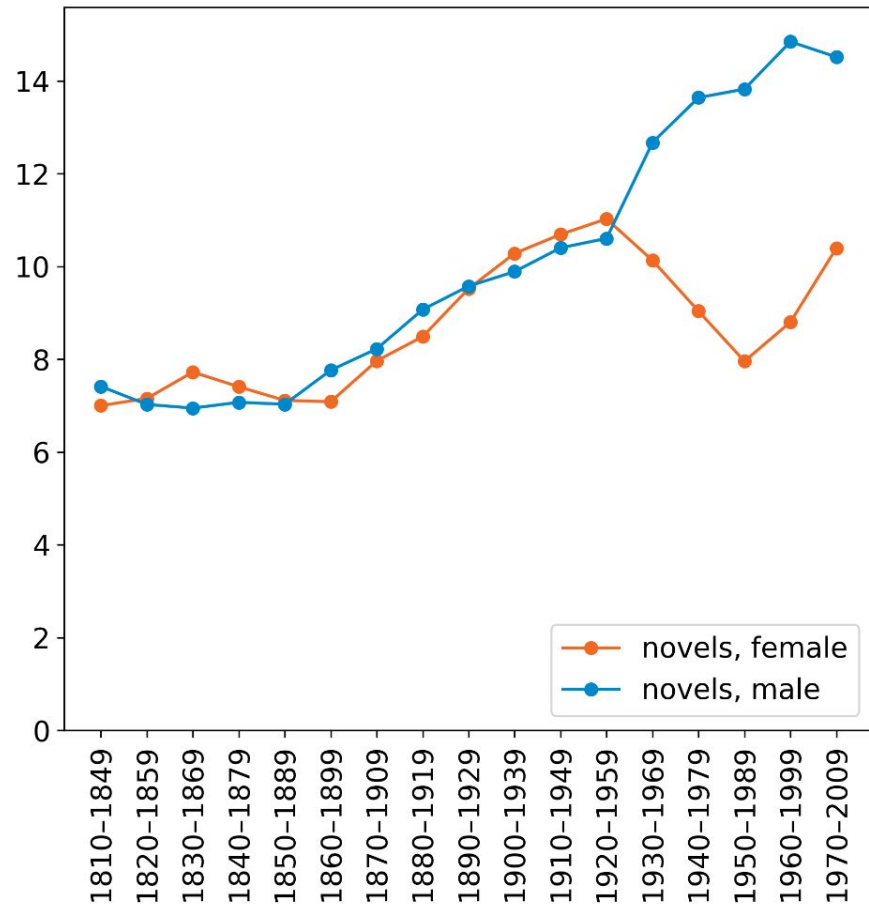
Significance of differences in time



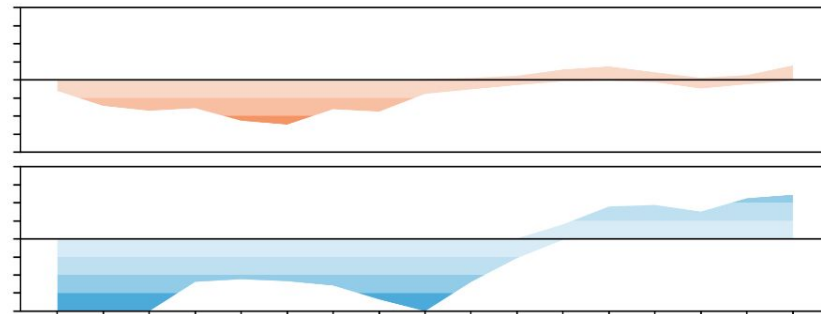
Significance in comparison with other categories



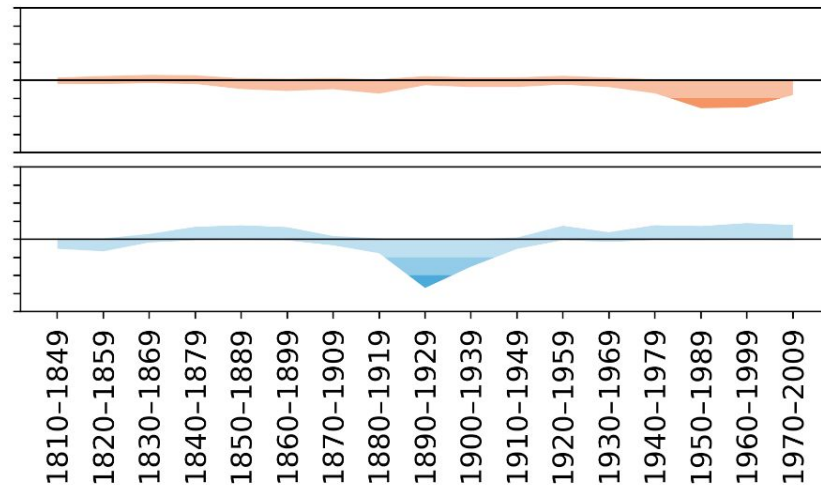
Hapaxes in subcorpora with 79 tokens



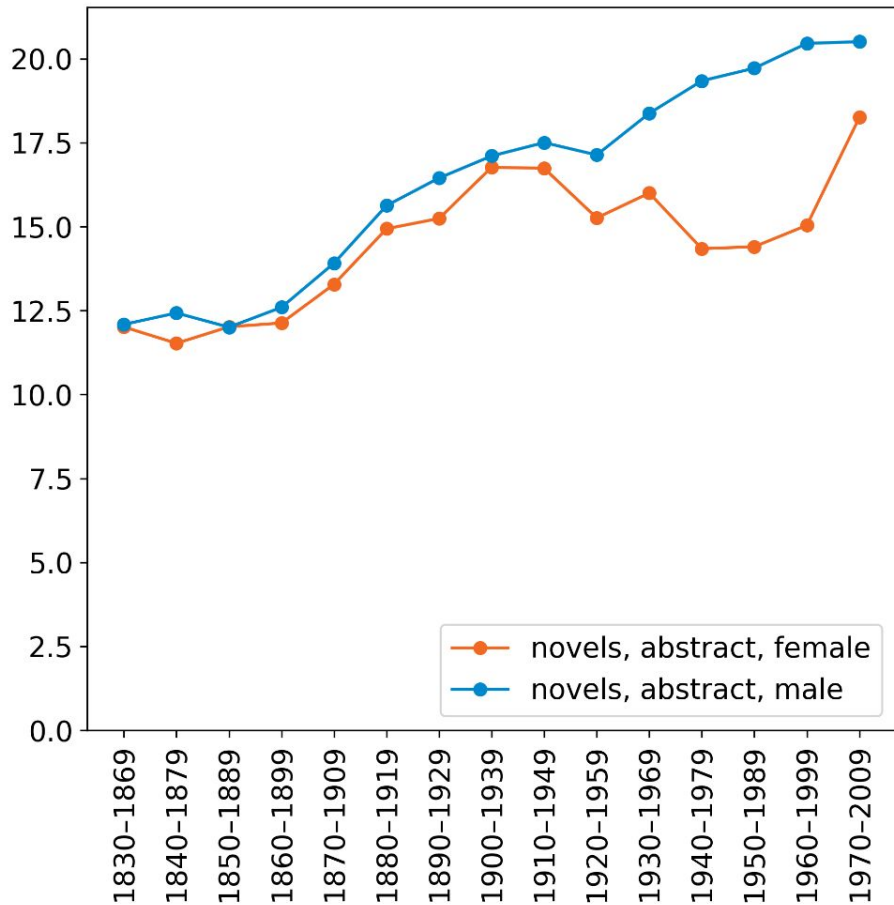
Significance of differences in time



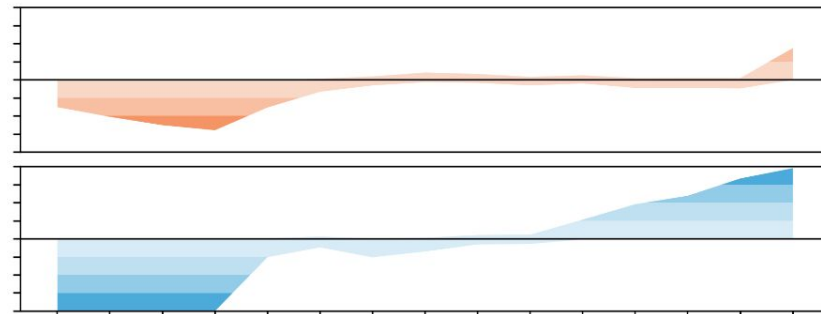
Significance in comparison with other categories



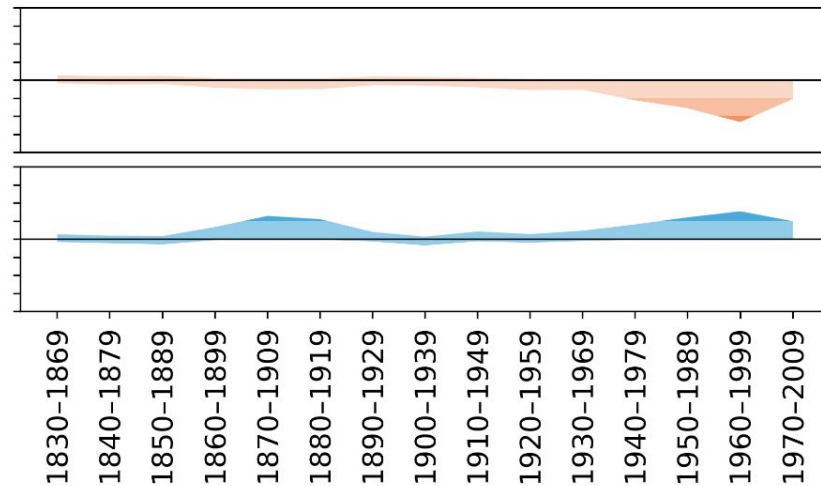
Types in subcorpora with 40 tokens



Significance of differences in time

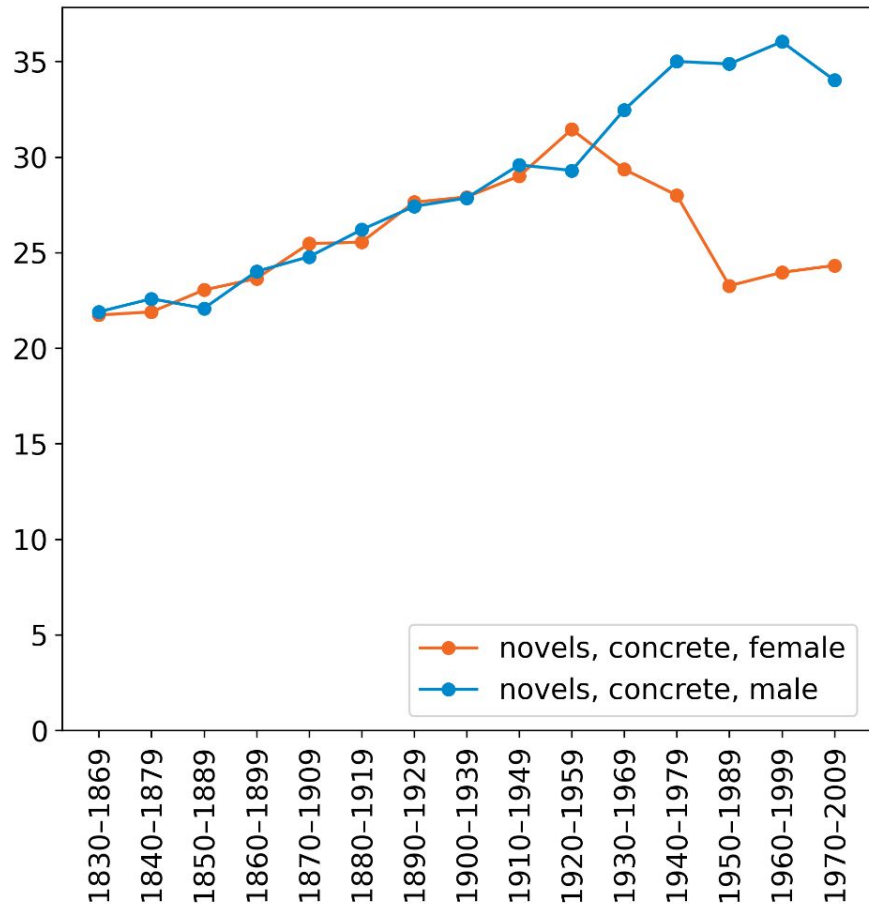


Significance in comparison with other categories

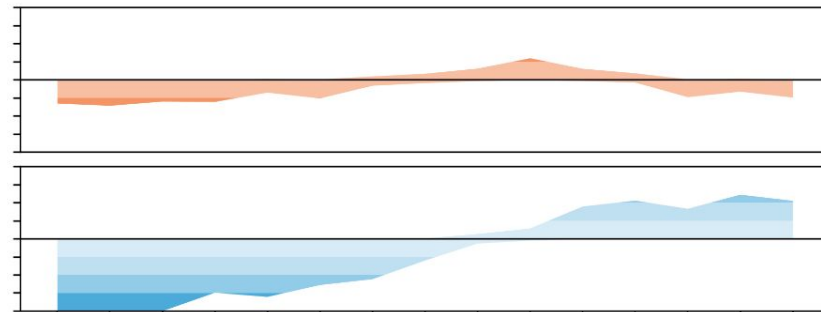




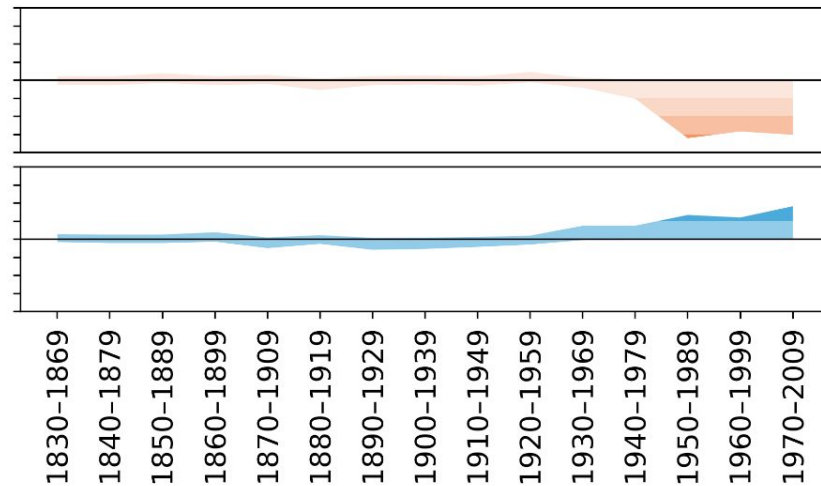
Types in subcorpora with 196 tokens



Significance of differences in time



Significance in comparison with other categories



# Type frequency

---

- Increase in productivity over time for both genders
  - Both abstract (slightly led by men?) and concrete types
  - Levelling in the later 20thC, dip in women's productivity
- But what *kind* of types are involved?
- Do men and women use the cx in different semantic areas?
- We examine these questions using ***distributional semantics***

## Analysis 2: semantic analysis

# Semantic analysis

---

- Same issues as type frequency: hard to compare and make semantic generalizations from uneven samples
- Types may vary considerably between matched-size random samples, so no sample can be representative
- Problem = we cannot average over individual types
  - But we can average over type counts
  - We just need to add a semantic dimension to type counts
  - We use distributional semantics to achieve this

# Distributional semantics

— — —

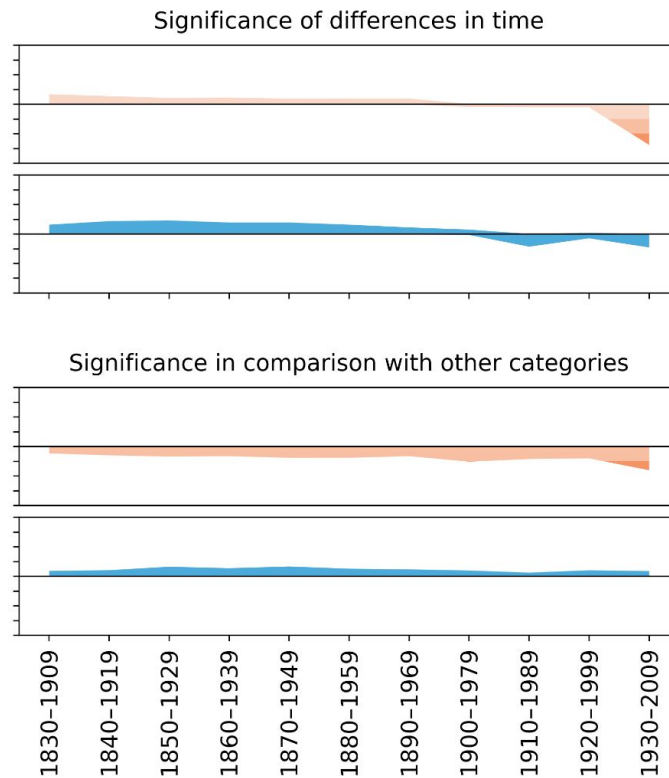
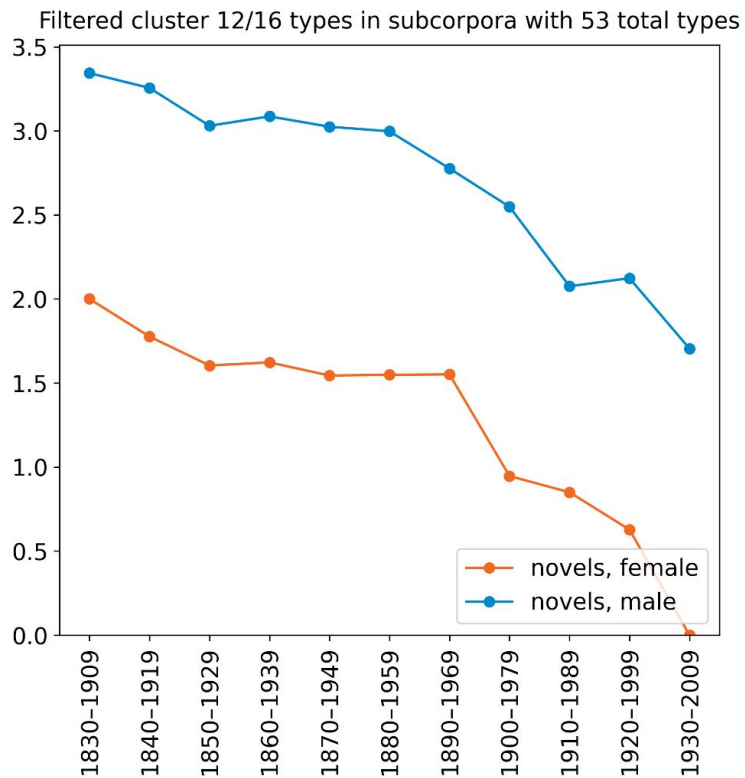
“You shall know a word by  
the company it keeps”  
Firth (1957: 11)

- Aim = capturing word meaning through lexical collocates in large text corpora
  - Semantically similar words are expected to have the same collocates
  - Semantic similarity is approximated by similarity in distribution
- DSM built with word2vec (SkipGram, cf. Mikolov et al. 2013), using gensim
- Trained on the whole COHA, context window  $\pm 2$  words

# Semantic analysis

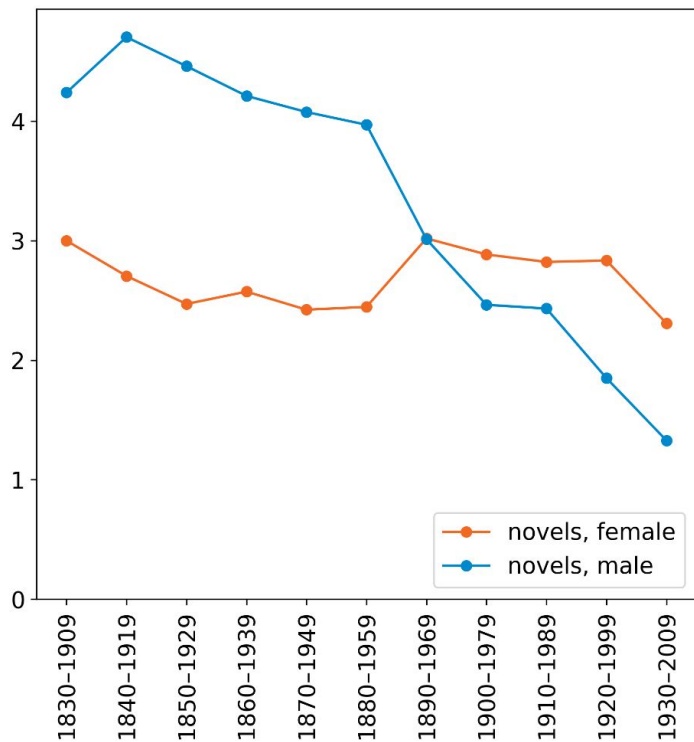
— — —

- We use this DSM to sort the types attested in the *way*-construction into semantic classes
  - Pairwise semantic similarity between types extracted from the DSM and submitted to cluster analysis (PAM)
  - 16-cluster solution: most clusters correspond to a clearly identifiable semantic category, with outliers
- Semantic classification manually corrected by removing the outliers
- Allows us to calculate type proportions for each semantic class and compare genders

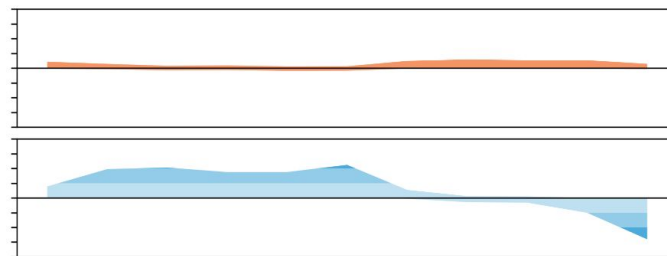


**Destruction:** blast burst bust crack rend rip smash tear

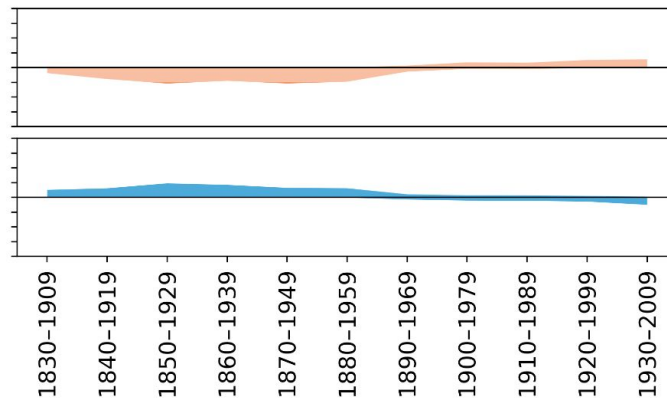
Filtered cluster 11/16 types in subcorpora with 53 total types



Significance of differences in time



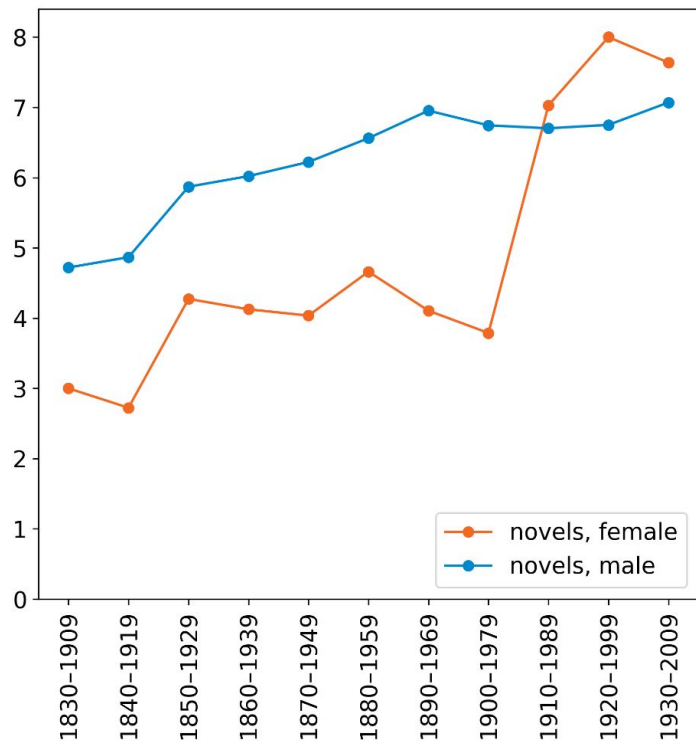
Significance in comparison with other categories



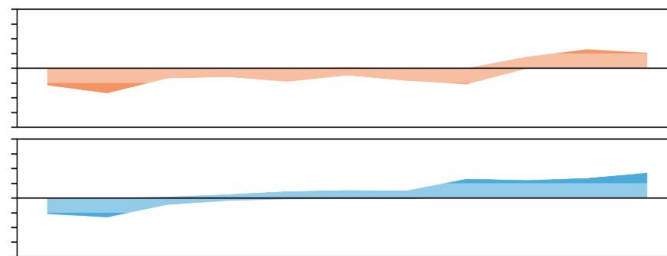
**Cutting/burning:** blaze burn cleave cleft etch furrow gnaw grate melt pierce ruffle scorch sear simmer smite sting



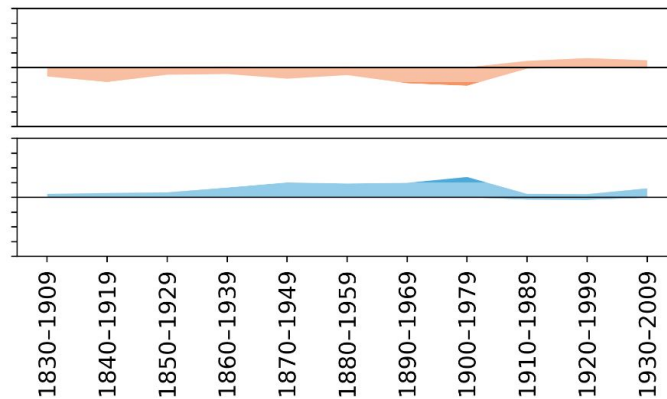
Filtered cluster 8/16 types in subcorpora with 53 total types



Significance of differences in time

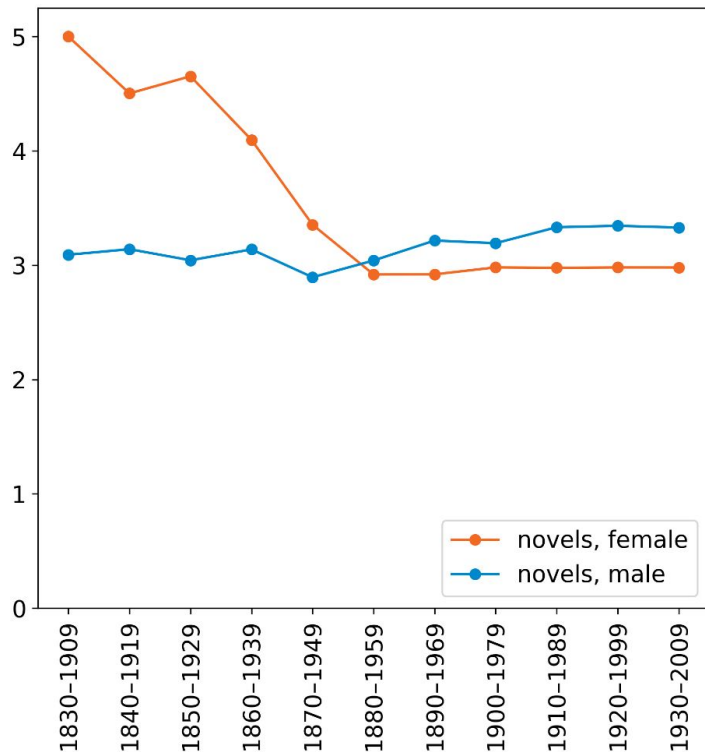


Significance in comparison with other categories

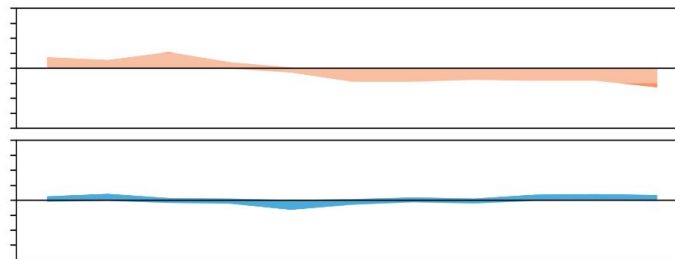


**Forceful contact, hollowing out:** burrow claw cling clutch cram dig finger fumble grind grope kick lick massage  
nose paw peck pound rattle scrabble scratch sniff swab thrash thumb

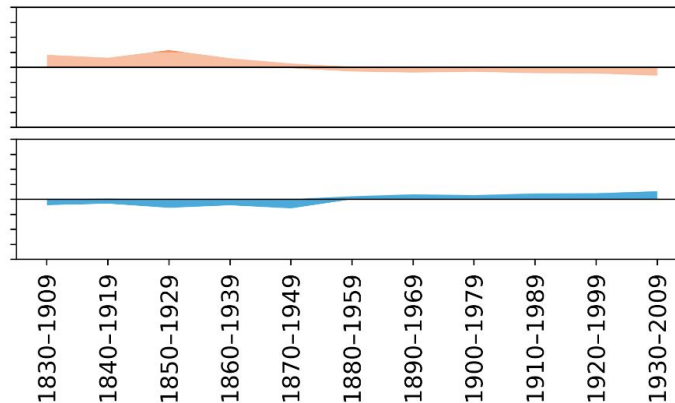
Filtered cluster 1/16 types in subcorpora with 53 total types



Significance of differences in time

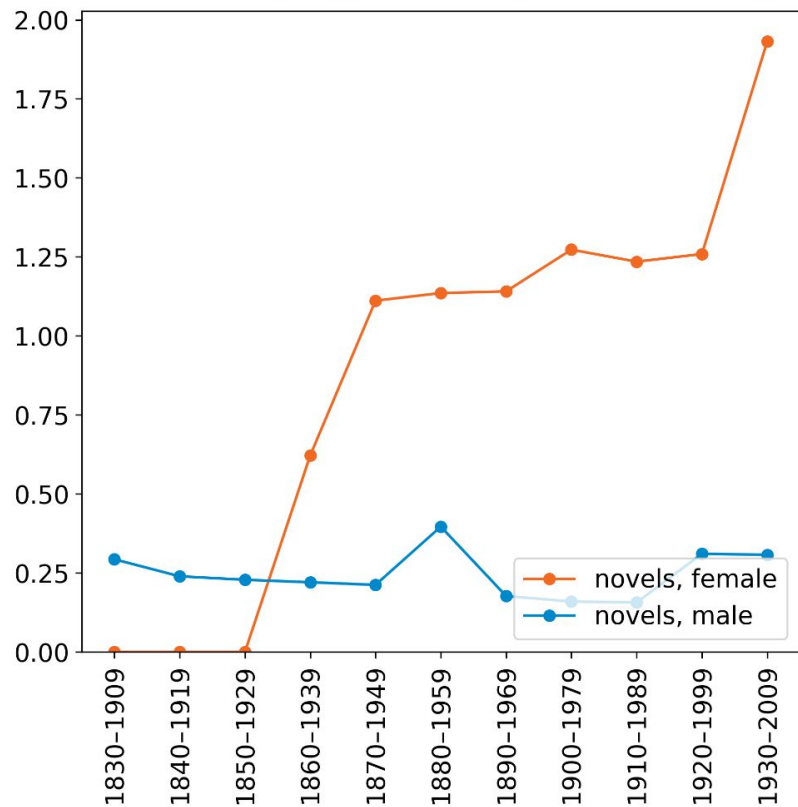


Significance in comparison with other categories

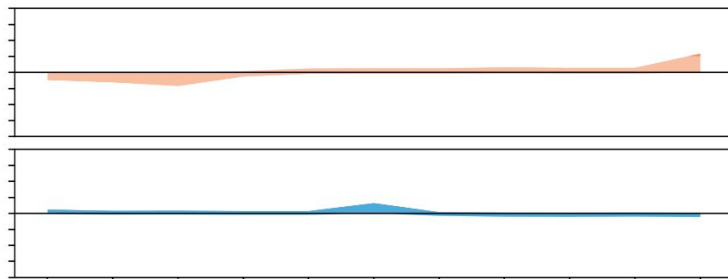


**Striving/fighting:** battle conquer dispute fend fight manage struggle work wrestle

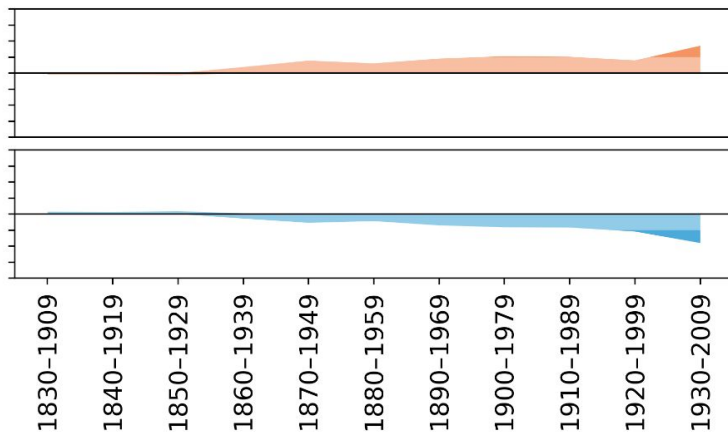
Filtered cluster 5/16 types in subcorpora with 53 total types



Significance of differences in time

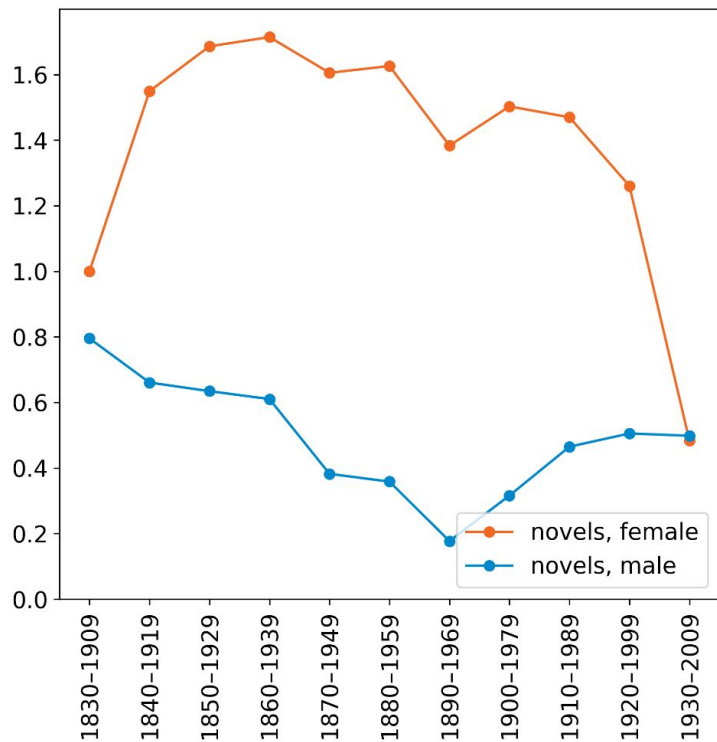


Significance in comparison with other categories

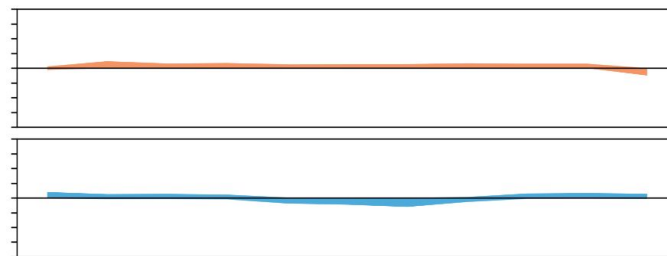


**Manner of speaking:** bawl bellow lisp mumble stammer whisper

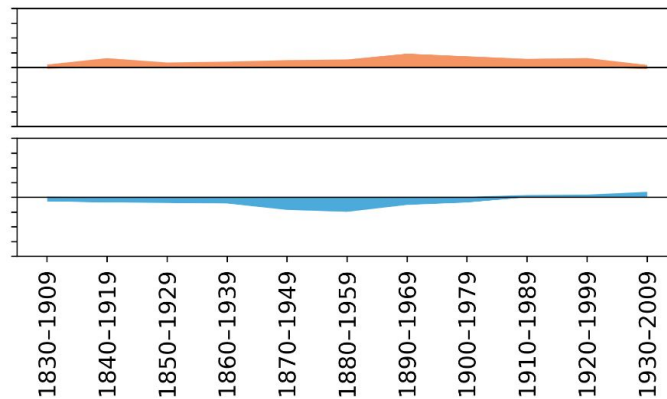
Filtered cluster 16/16 types in subcorpora with 53 total types



Significance of differences in time

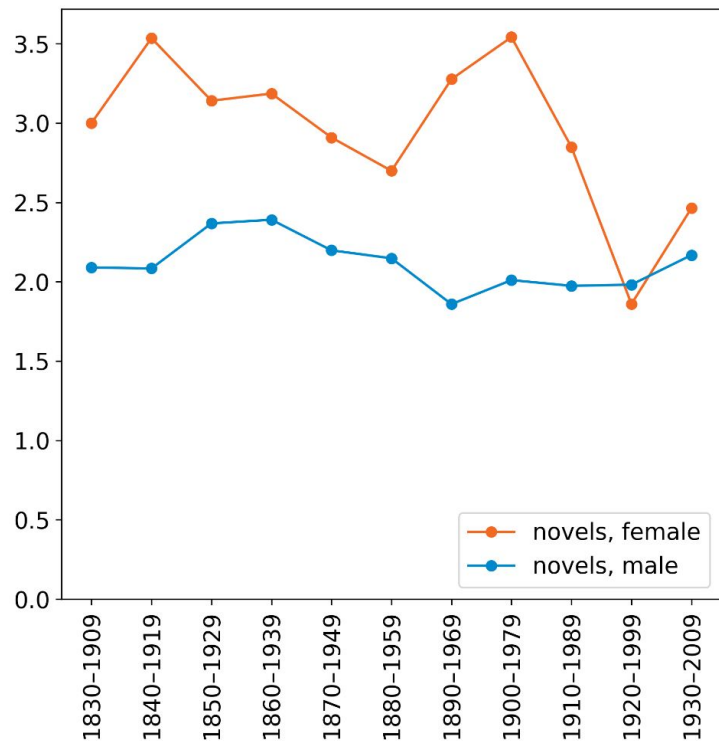


Significance in comparison with other categories

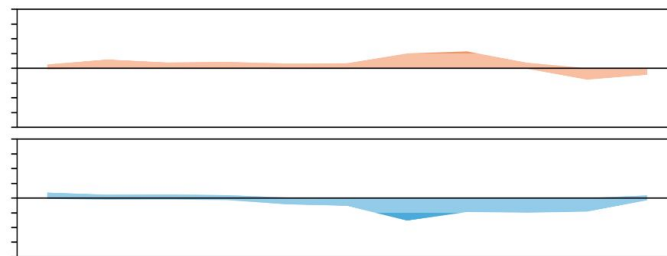


**Directing:** direct guide lead pilot steer

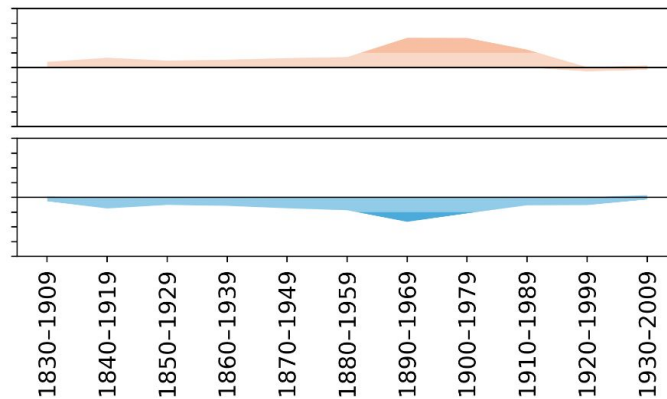
Filtered cluster 15/16 types in subcorpora with 53 total types



Significance of differences in time

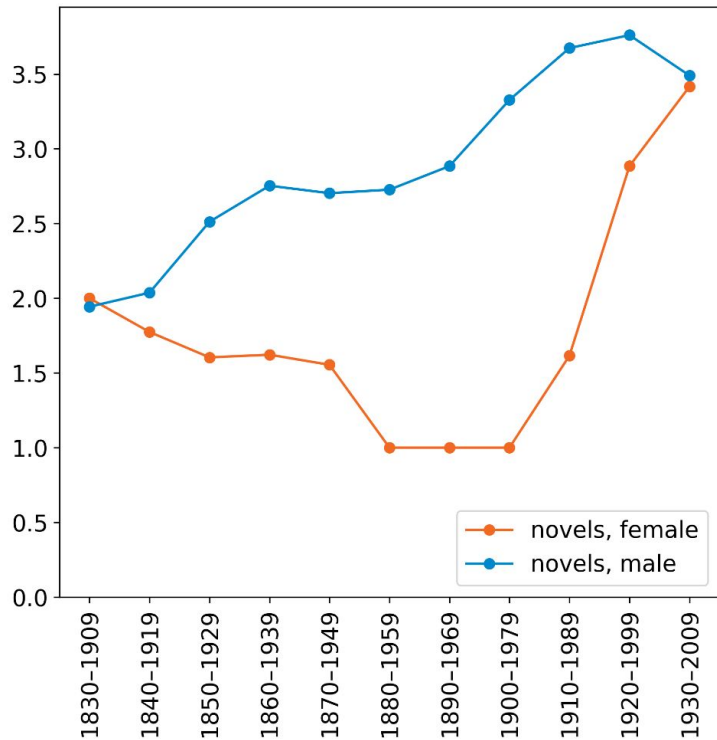


Significance in comparison with other categories

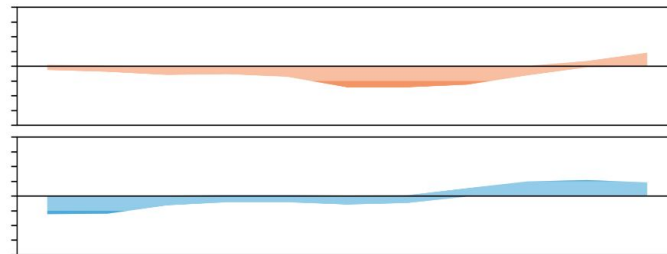


**Cognition:** daydream dream feel guess plan sleepwalk think understand worry

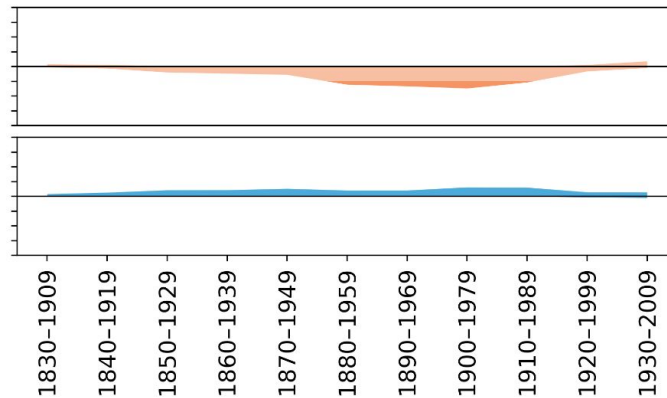
Filtered cluster 6/16 types in subcorpora with 53 total types



Significance of differences in time

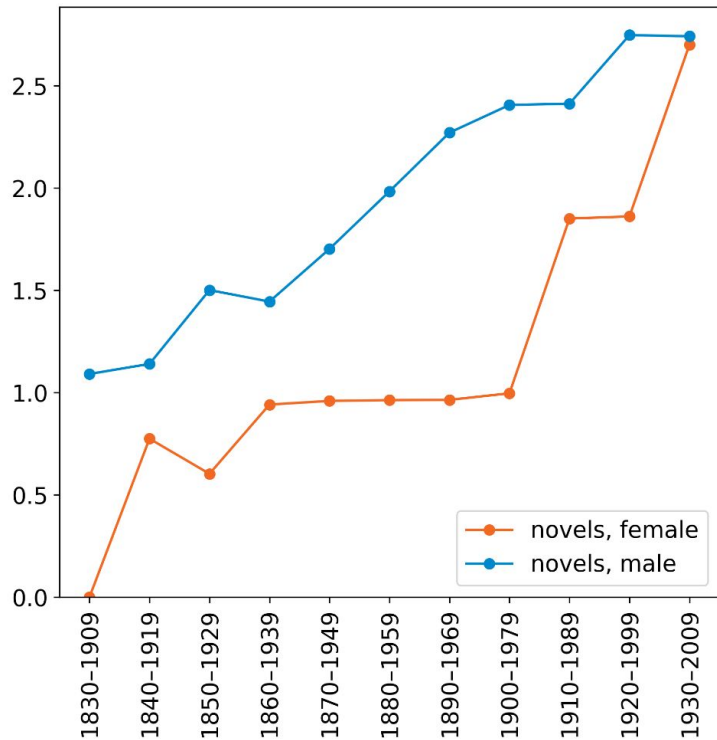


Significance in comparison with other categories

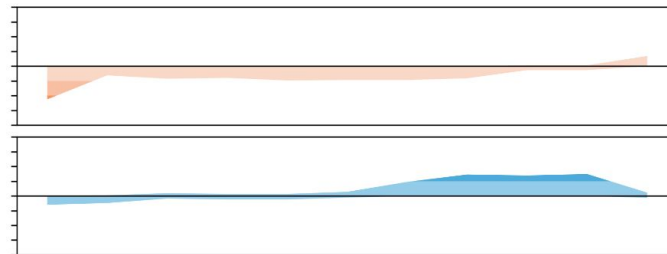


**Coercion/trickery:** beg bribe bully charm cheat coax connive fawn force forge scheme terrorize trick wheedle

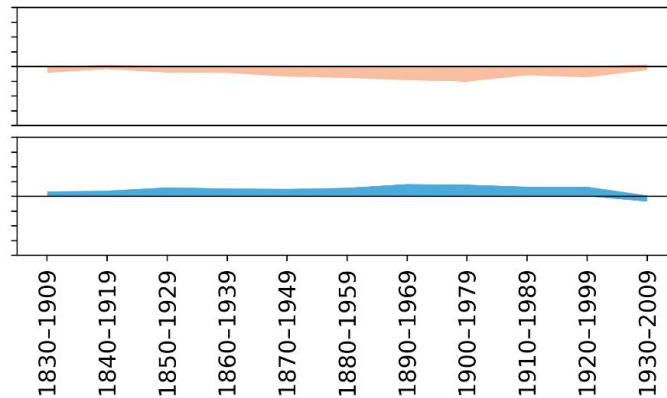
Filtered cluster 3/16 types in subcorpora with 53 total types



Significance of differences in time

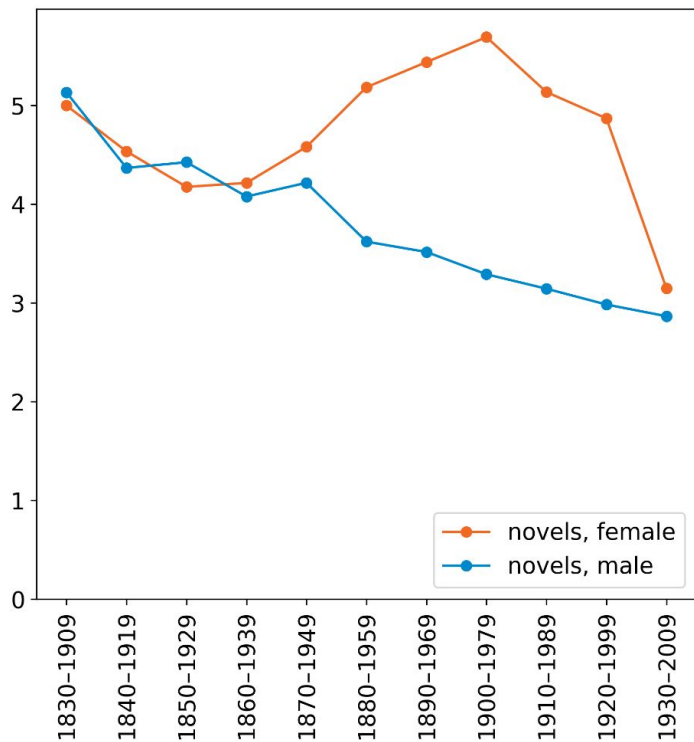


Significance in comparison with other categories

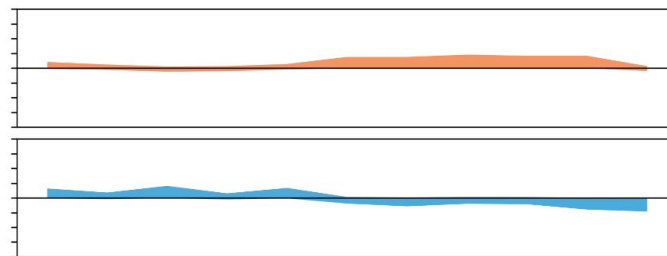


**Communication:** argue chat explain joke kid lie negotiate read sing talk write

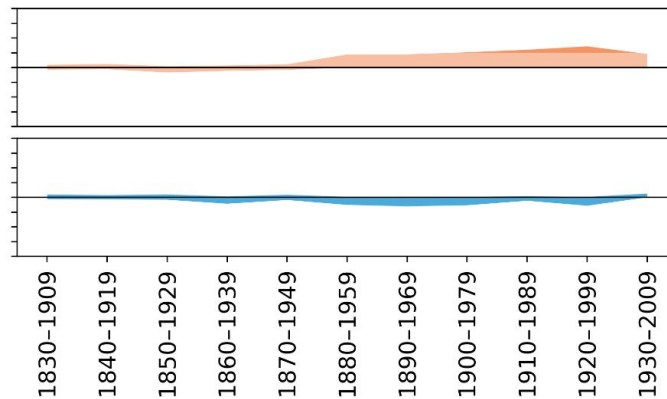
Filtered cluster 10/16 types in subcorpora with 53 total types



Significance of differences in time



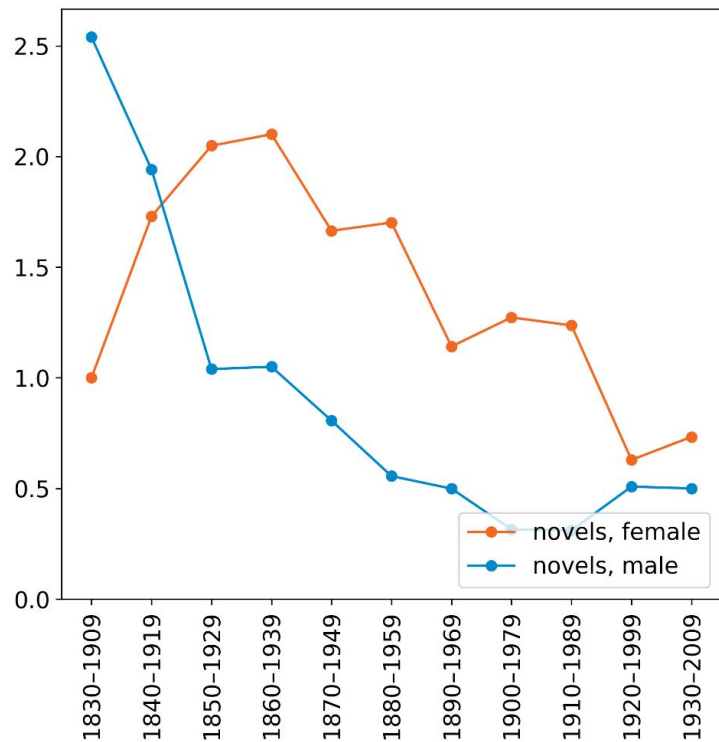
Significance in comparison with other categories



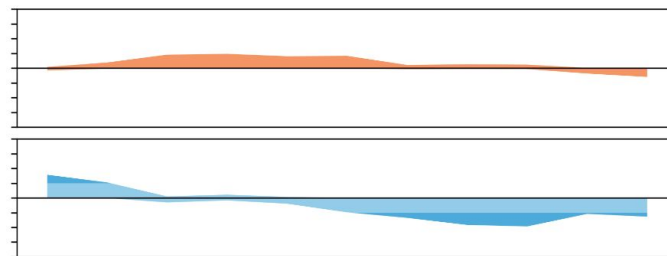
**Damaging:** break carve chip chop cut hew saw shear shoot slash slice strike whittle



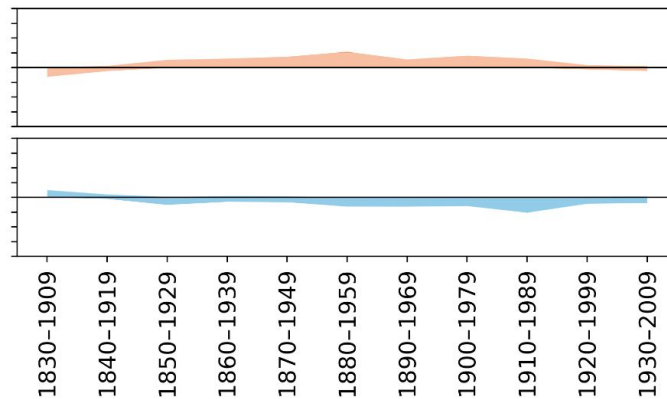
Filtered cluster 9/16 types in subcorpora with 53 total types



Significance of differences in time



Significance in comparison with other categories



**Searching:** brave eke explore ferret leak marshal pillage probe scent sort trace track

# Discussion

# Summary of results

---

- Gender differences difficult to interpret
- Differences in overall type frequency only found in the late 20thC (men more productive, women less)
- In the semantic analysis, some classes are more in line with the original “path-creation” meaning than others  
e.g. cutting, destruction vs. speaking, trickery
- Men tend to favour the former classes
- The opposite tends to be true for women, with exceptions

# Conclusions

— — —

- Both genders participate in the changes in the productivity of the construction
- But they do so in different ways
- Possible explanations / confounds to be explored:
  - Changes in men and women's usage of different verb types over time?
  - Changes in the novel genres sampled in the COHA over time and between genders?
  - *Way*-construction = marker of creativity, hence more productive in some genres?
  - Competition with another construction favoured by women?

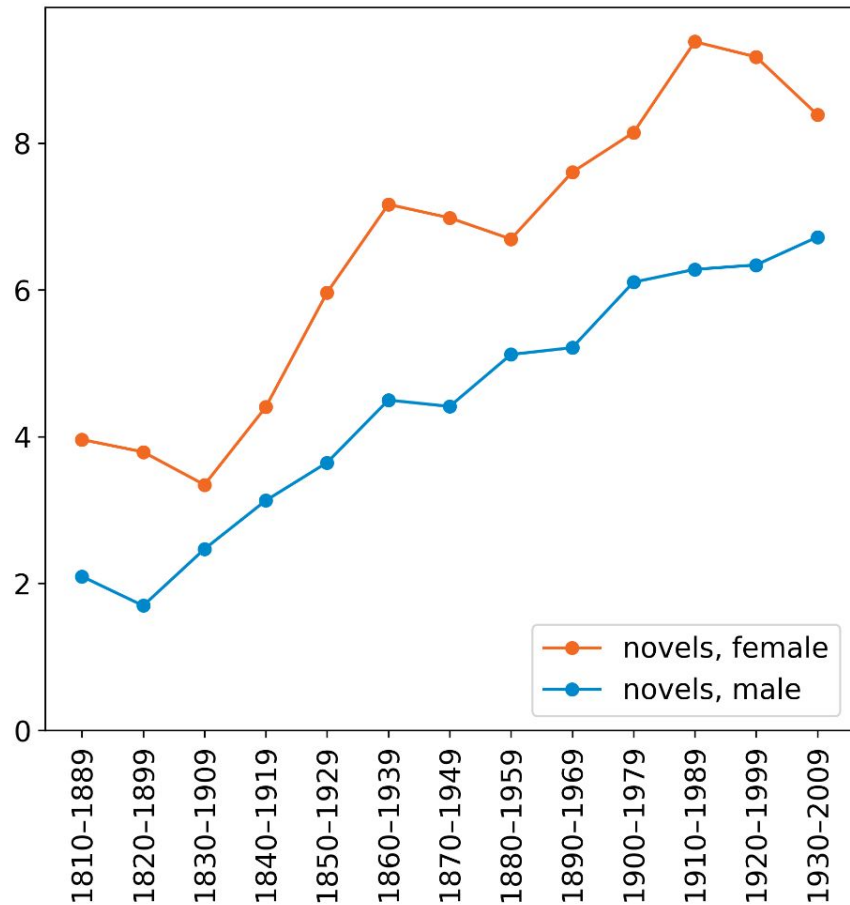
# References

— — —

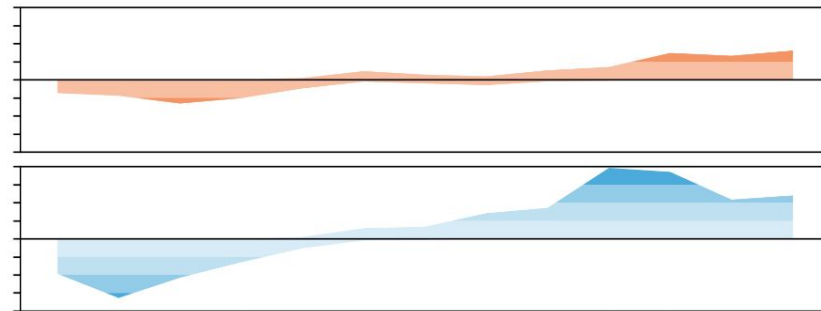
- Israel, Michael. 1996. The way constructions grow. In Adele E. Goldberg (ed.), *Conceptual structure, discourse and language*, 217–230. Stanford, CA: CSLI Publications.
- Öhman, Emily, Tanja Säily & Mikko Laitinen. 2019. Towards the inevitable demise of everybody? A multifactorial analysis of *-one/-body/-man* variation in indefinite pronouns in historical American English. 40th Annual Conference of the International Computer Archive of Modern and Medieval English (ICAME 40), Neuchâtel, Switzerland, June 2019. [https://tanasaily.fi/talks/icame40\\_ohman\\_et\\_al\\_2019.pdf](https://tanasaily.fi/talks/icame40_ohman_et_al_2019.pdf)
- Nevalainen, Terttu & Helena Ramoulin-Brunberg. 2003. *Historical Sociolinguistics: Language change in Tudor and Stuart England*. London: Pearson Education.
- 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.
- Perek, Florent (2020). Productivity and schematicity in constructional change. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and Links in the Network: Advances in Diachronic Construction Grammar* (pp. 141–166). Amsterdam: John Benjamins.
- 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.
- Säily, Tanja. 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1): 119–141.
- 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*. Amsterdam: John Benjamins.
- Säily, Tanja & Turo Vartiainen. Forthcoming. Historical linguistics. In Michaela Mahlberg & Gavin Brookes (eds.), *Bloomsbury handbook of corpus linguistics*. Bloomsbury.
- Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

**Bonus slides**

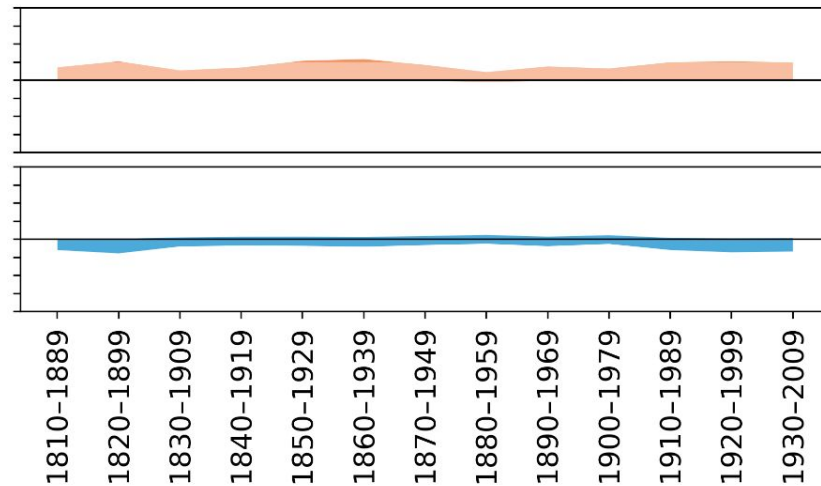
Incidental types in subcorpora with 58 total types



Significance of differences in time



Significance in comparison with other categories



# Why the gender gap?

---

It seems to be a particularly robust phenomenon;  
none of these can explain it (alone):

- a small number of authors with a peculiar style
- a small number of words used exclusively by men
- random chance, specific choice of time periods, etc.

**Very preliminary** indicators pointing in this direction:

- not so much **men starting** to use new verbs with *way*?
- but maybe **women stopping** to use some old verbs that men still keep using with *way*?





# Data

- Verb + Possessive determiner + *way* + preposition  
extracted from the corpus, manually filtered
- Annotated for sense: path-creation, manner, incidental
- Annotated for concrete vs. abstract motion
- 7,455 tokens (2,120 from women vs. 5,335 from men)

	Total	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
women	2120	0	26	32	21	119	121	128	98	175	150	181	124	86	75	58	55	51	146	115	359
men	5335	5	57	321	381	328	325	250	295	354	431	361	414	185	206	217	233	312	290	99	271