

CHANGING STYLES OF LETTER-WRITING? EVIDENCE FROM A POS-TAGGED CORPUS OF 18TH-CENTURY LETTERS

Tanja Säily, Turo Vartiainen, Harri Siirtola and Terttu Nevalainen



INTRODUCTION

- Personal letters: “speech-like” genre (Culpeper & Kytö 2010: 17), provide access to everyday language use in the past
- How has the genre changed over time? Sociolinguistic variation?
- Biber & Finegan (1997), Biber (2001): multidimensional studies of a number of linguistic features in the ARCHER corpus (17th–20th centuries)
 - Letter genre becomes less informational and more personally involved over time



FEATURES OF INVOLVED vs. INFORMATIONAL TEXTS IN 18TH-CENTURY ENGLISH

Dimension 1

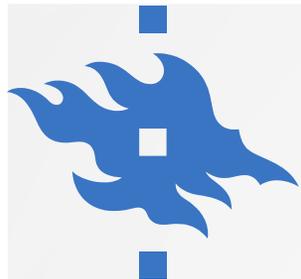
Positive features

first-person pronouns	0.92
second-person pronouns	0.80
present tense verbs	0.79
possibility modals	0.68
<i>that</i> deletion	0.66
mental verbs	0.65
prediction modals	0.59
general emphatics	0.55
conditional subordination	0.55
<i>do</i> as pro-verb	0.42
necessity modals	0.40
<i>be</i> as main verb	0.36
indefinite pronouns	0.36
speech act verbs	0.30

Negative features

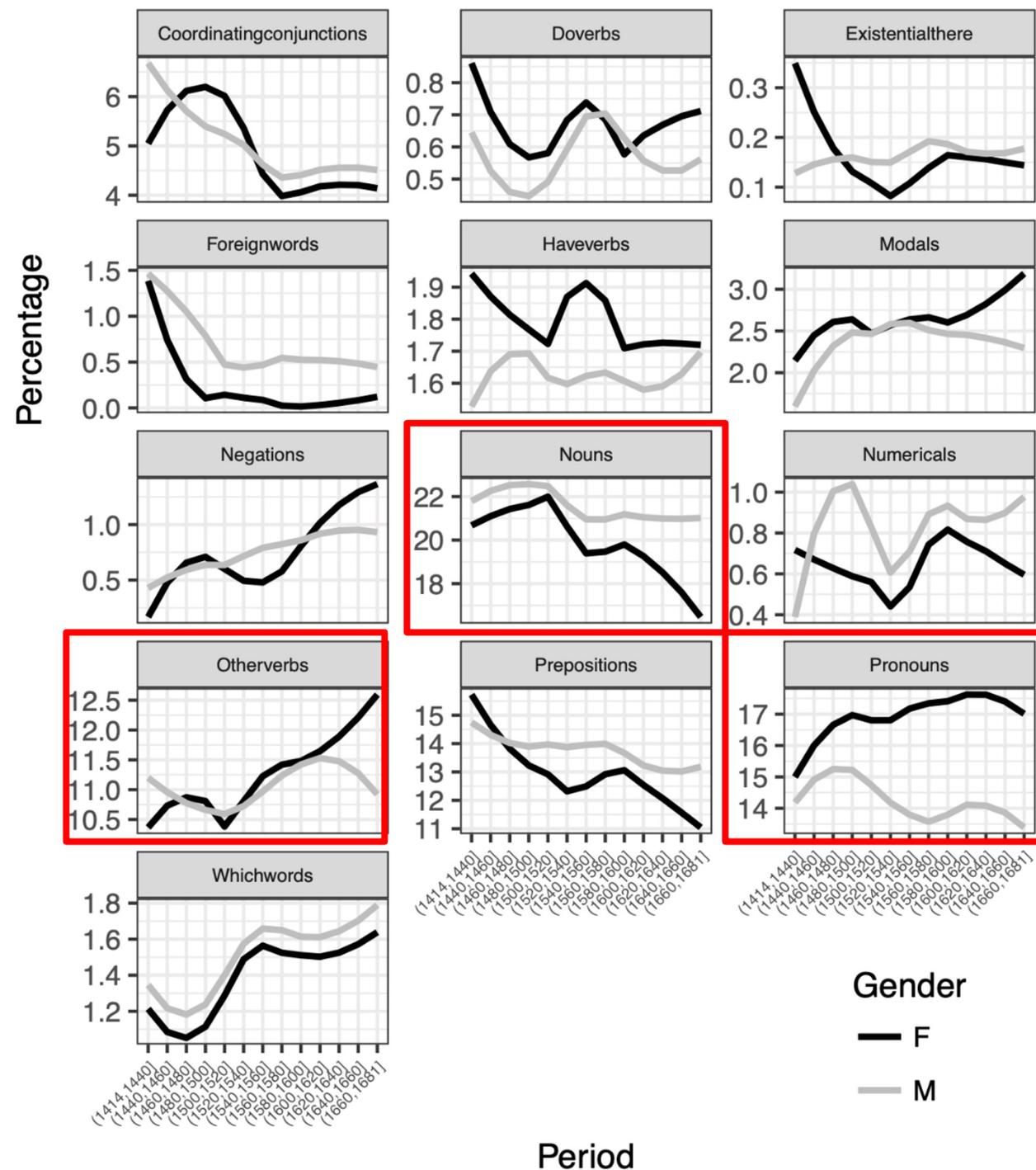
prepositions	-0.64
agentless passives	-0.64
nouns	-0.62
reduced passive postnominal clauses	-0.57
word length	-0.55
past tense verbs	-0.51
<i>by</i> passives	-0.40
type / token ratio	-0.36

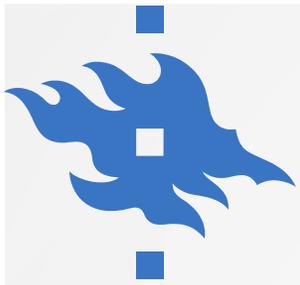
Biber, Douglas. 2001. Dimensions of variation among 18th-century speech-based and written registers. *Towards a history of English as a history of genres* (Anglistische Forschungen 298), ed. by Hans-Jürgen Diller & Manfred Görlach, 89-109. Heidelberg: Universitätsverlag C. Winter. (p. 98)



TRENDS BASED ON POS RATIOS (1)

- Säily et al. (2011, 2017): part-of-speech ratios in 15th–17th-century correspondence (PCEEC)
 - Proportion of nouns decreased, lexical verbs increased → **colloquialization?**
 - **Gendered styles** of writing: women consistently used fewer nouns and more personal pronouns than men → leaders of change towards involved style?





TRENDS BASED ON POS RATIOS (2)

- Vartiainen et al. (2013): first- & second-person pronouns, 17th–18th c. (CEEC)
 - Better match for involved style than all personal pronouns
 - Support for gendered styles, but also audience design (Bell 1984)
 - Gender differences decreased/disappeared in the 18th century → **shared writing style** among upper classes?
 - Similar result for nominal suffix *-ity* by Säily (2016, 2018)
 - Societal factors: women’s increased access to education, “feminization” of culture (McIntosh 2008: 231)
 - More research needed!



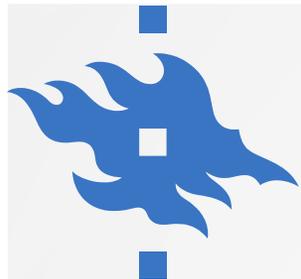
TAGGED CORPUS OF EARLY ENGLISH CORRESPONDENCE EXTENSION (TCEECE)

- Long 18th century: data chiefly from 1680–1800
- Based on published original-spelling editions sampled & digitized by team at the University of Helsinki
 - For this version, spelling standardized using VARD
- Compiled for historical sociolinguistics: metadata on letters, writers, recipients
- 2.2 million words, represents a **wide social spectrum** → enables more nuanced analyses than ARCHER
- **POS tagging**: CLAWS7 (Saario & Säily 2020)
 - Accuracy c. 94.5%



RESEARCH QUESTIONS

- Noun frequencies ~ **informational style**
 - Continued decline from PCEEC, supporting results from ARCHER?
 - Continued gender difference from PCEEC or shared style as implied by pronouns?
 - Influence of other social factors?
 - Social rank
 - Relationship between sender & recipient (FN = nuclear family, TC = close friends, T = other)
- Lexical verbs ~ **involved style**
 - Continued increase from PCEEC, supporting results from ARCHER?
 - Influence of social factors?



TOOL FOR EXPLORATION

- Developed by students taking the Data Science Project course at the University of Helsinki
 - Clients: Tanja Säily and Terttu Nevalainen
 - Students: Arthur Hughes, Hilma Holappa, Miia Rämö, Mikko Lipsanen, Pyörni Vartiainen
- Visual exploration of POS ratios over time, filtering by social category
 - POS tags can be grouped, as can years and social categories

Instructions

Line

Bar

Pronouns (I, me, you)



Select time range:

Selected period: 1680 - 1800

Select period length for grouping 40

Set name for the graph Pronouns (I, me, you)

Which lines are visible:

Line 1 Line 2 Line 3 Line 4 Line 5 Line 6 Line 7 Line 8 Line 9 Line 10

Apply selection

► Apply same POS tag selections for all lines

► Apply same attribute selections for all lines

▼ Line 1

Custom name Men

Sender sex

M

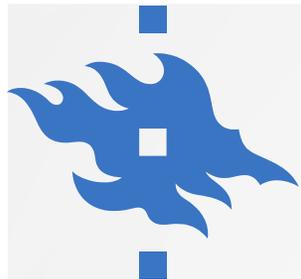
Sender rank

Regular

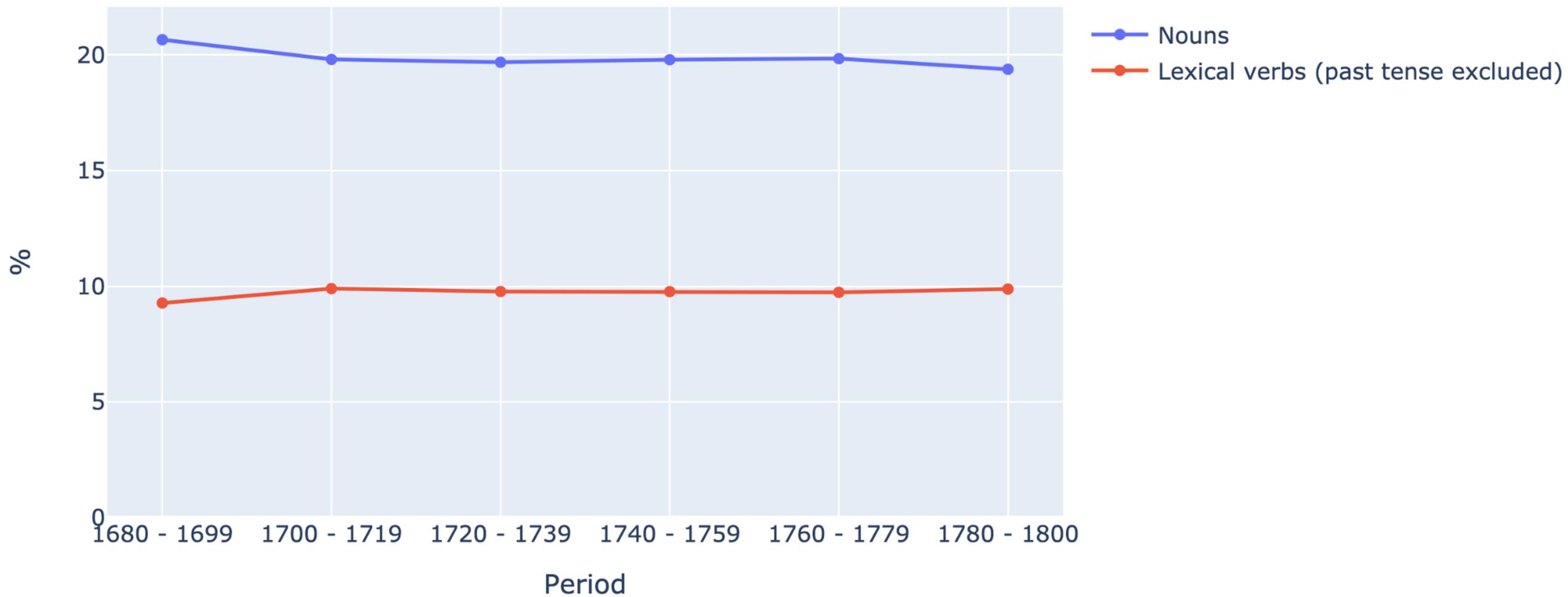
Royalty Nobility Gentry Clergy Professional Merchant Other

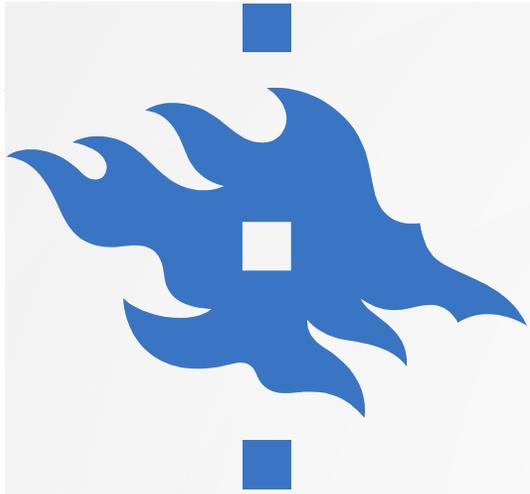
Relationship

Fine-grained



PILOT RESULTS

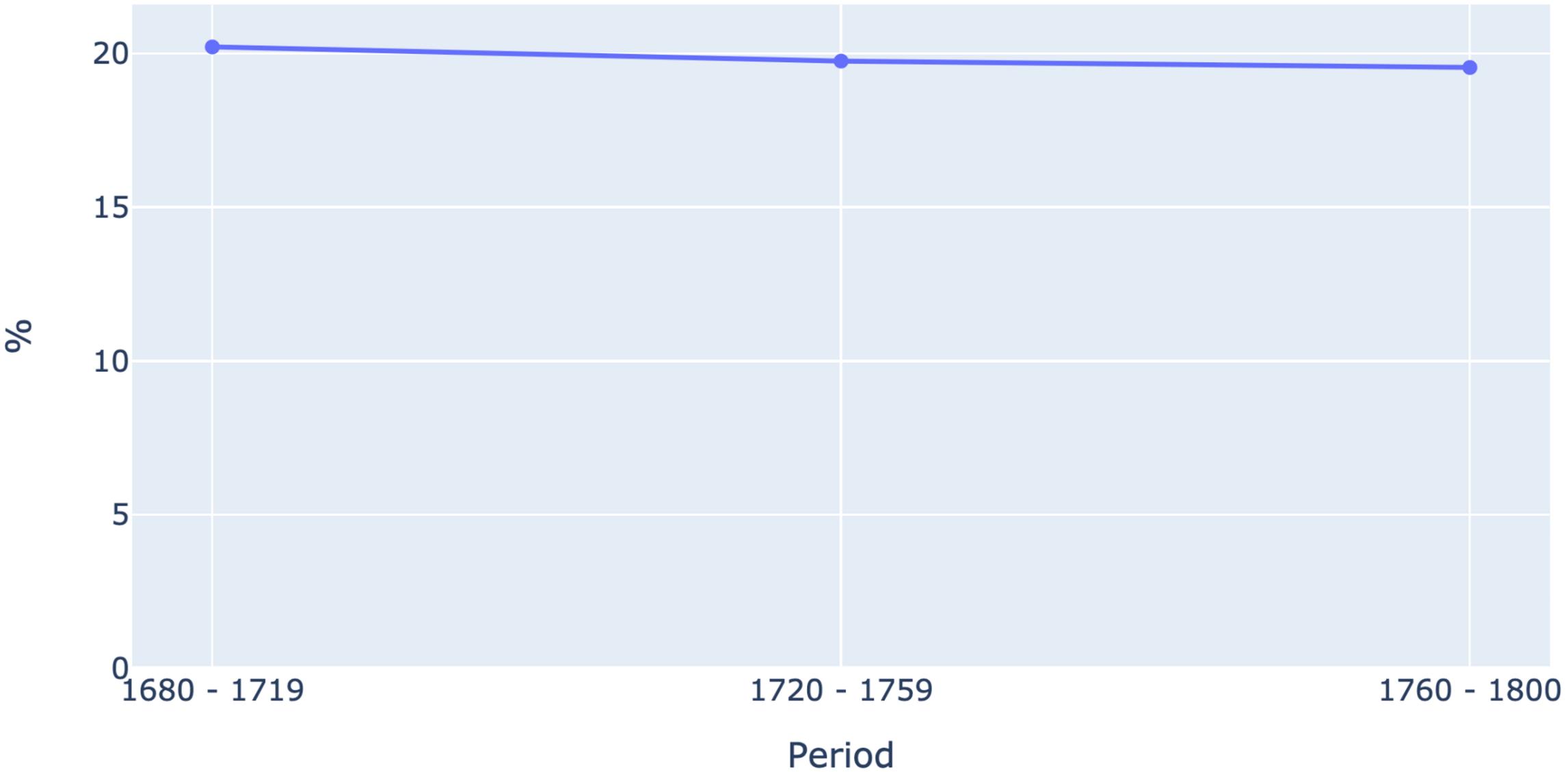




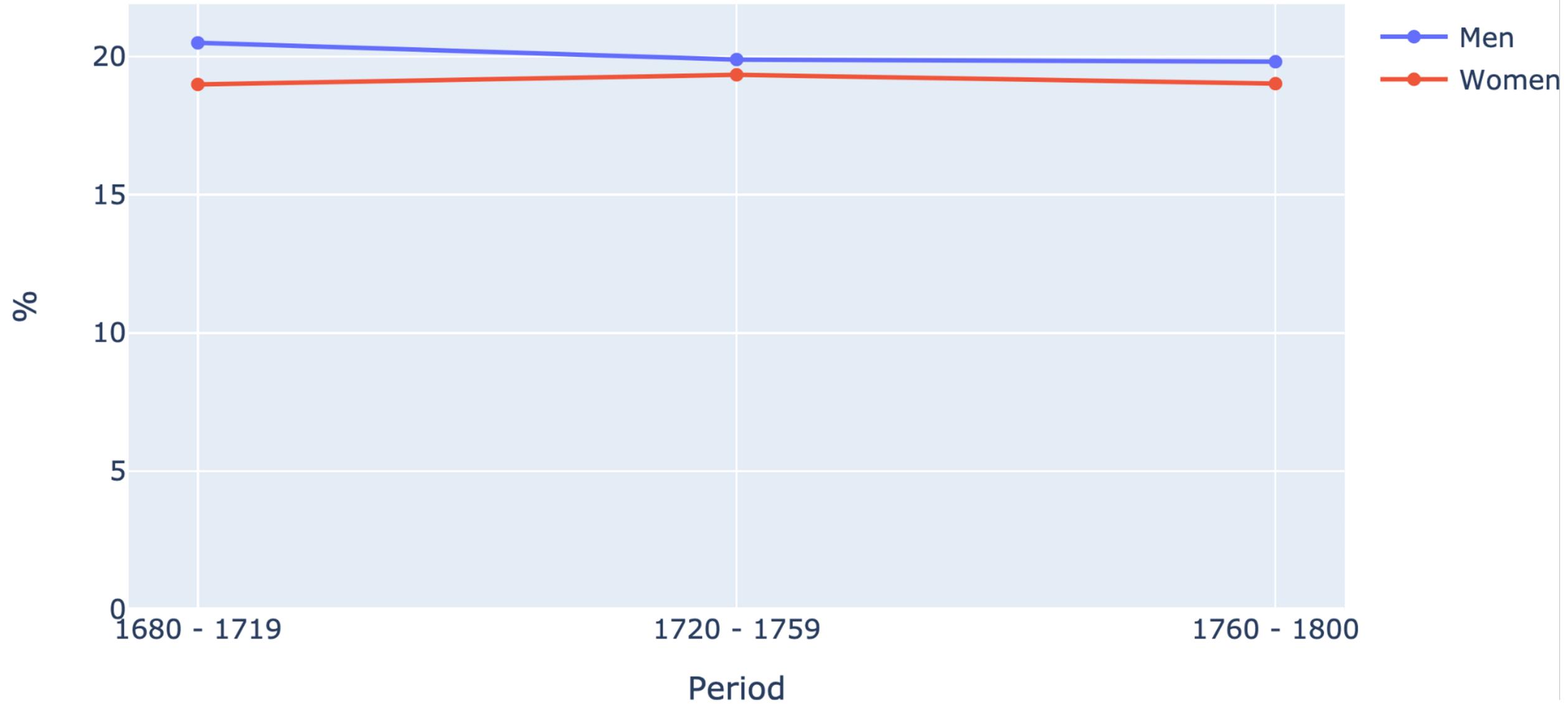
NOUNS

+ GENDER + RELATIONSHIP + SOCIAL RANK

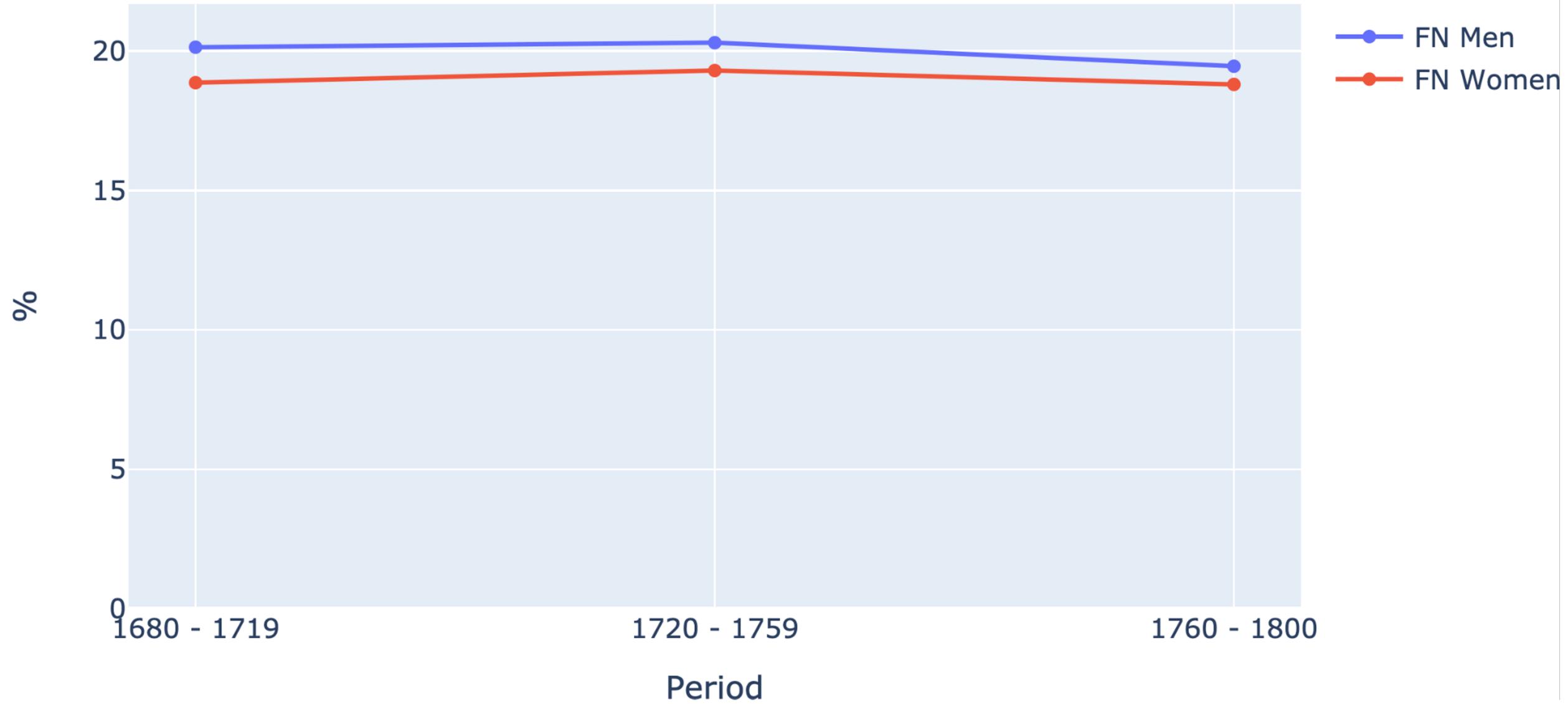
Nouns



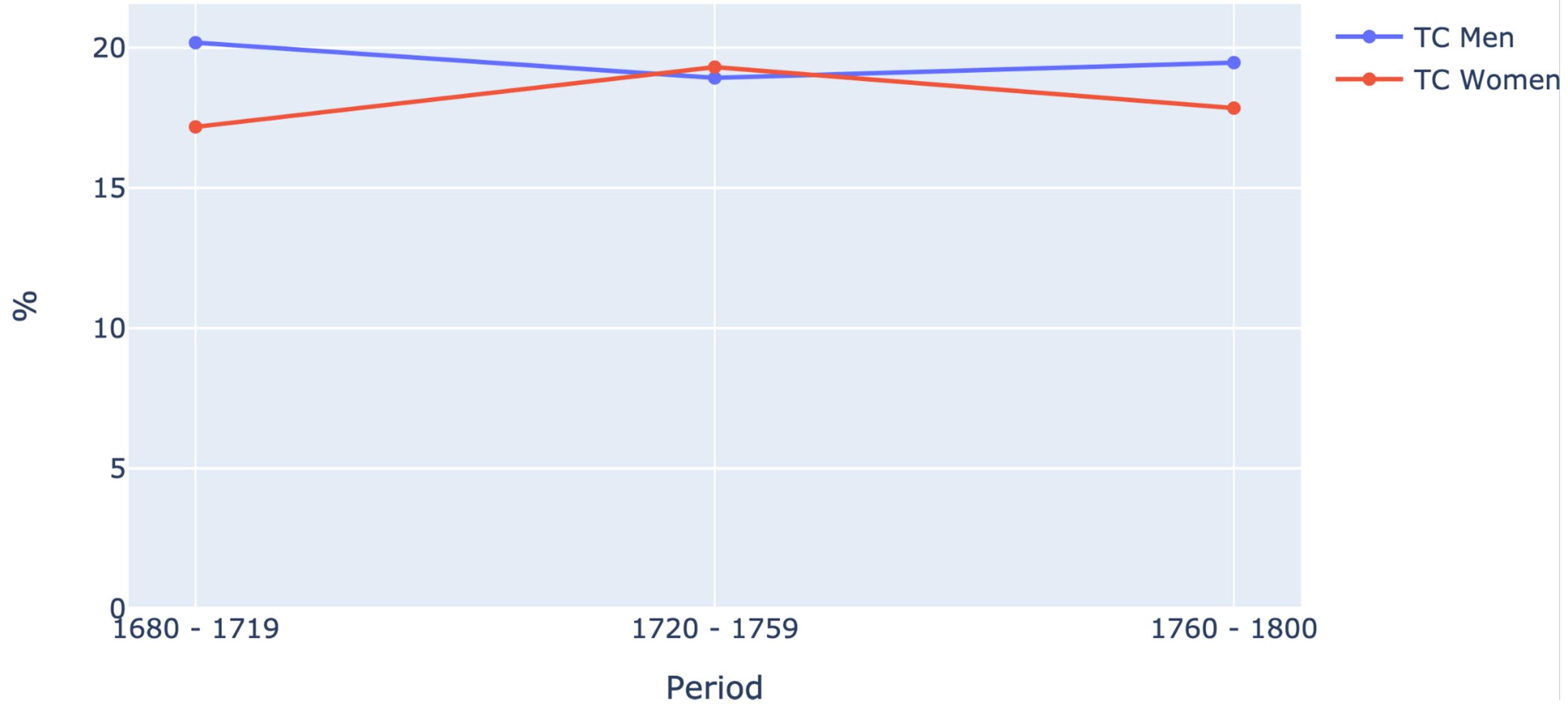
Nouns



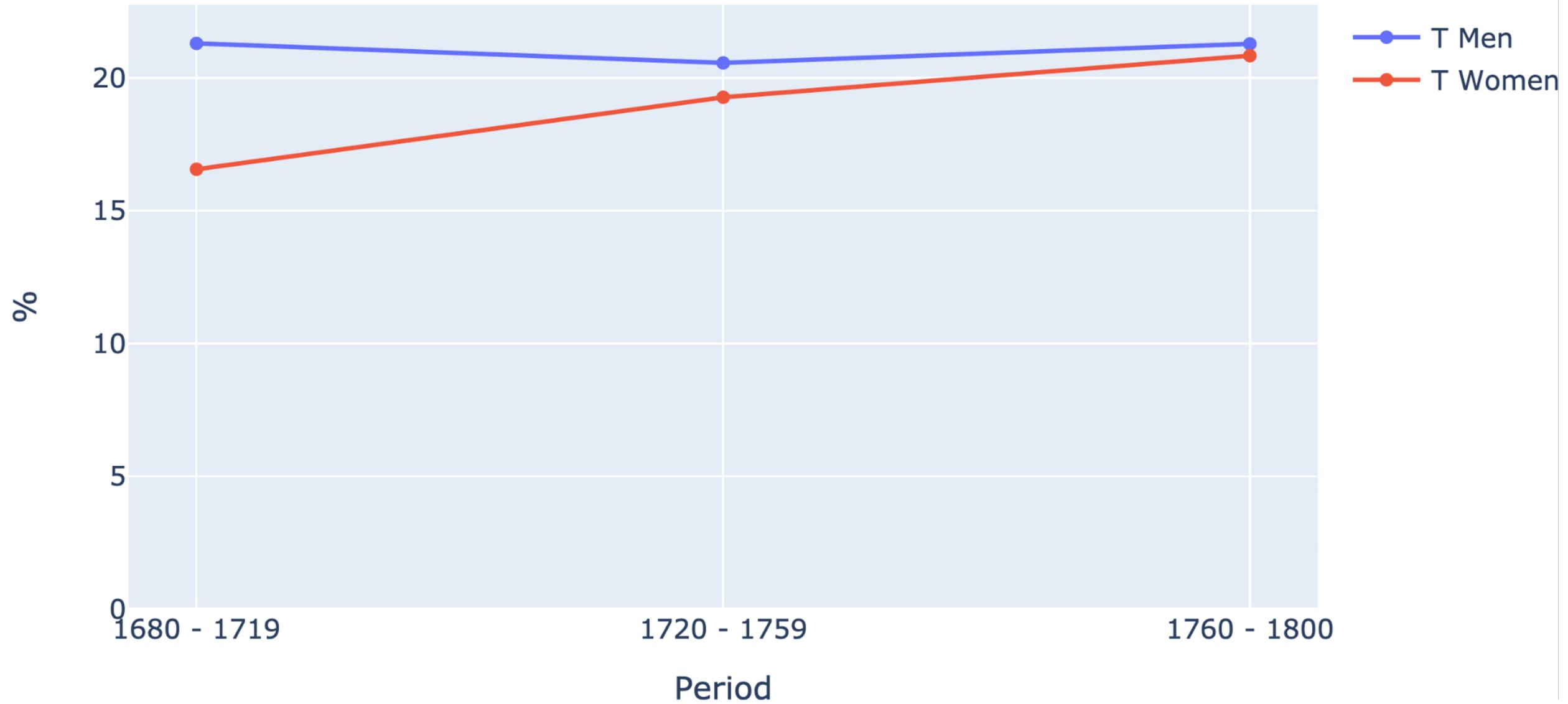
Nouns



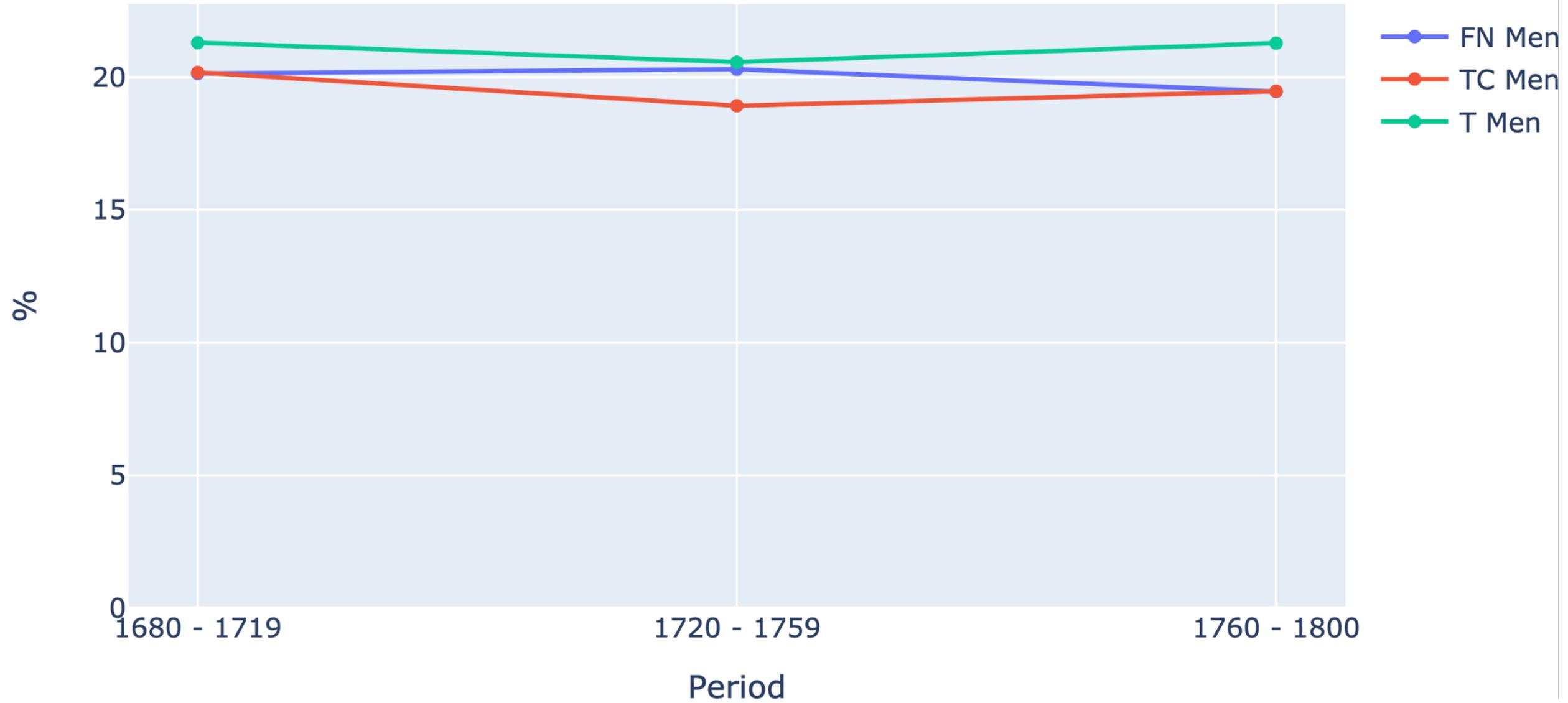
Nouns



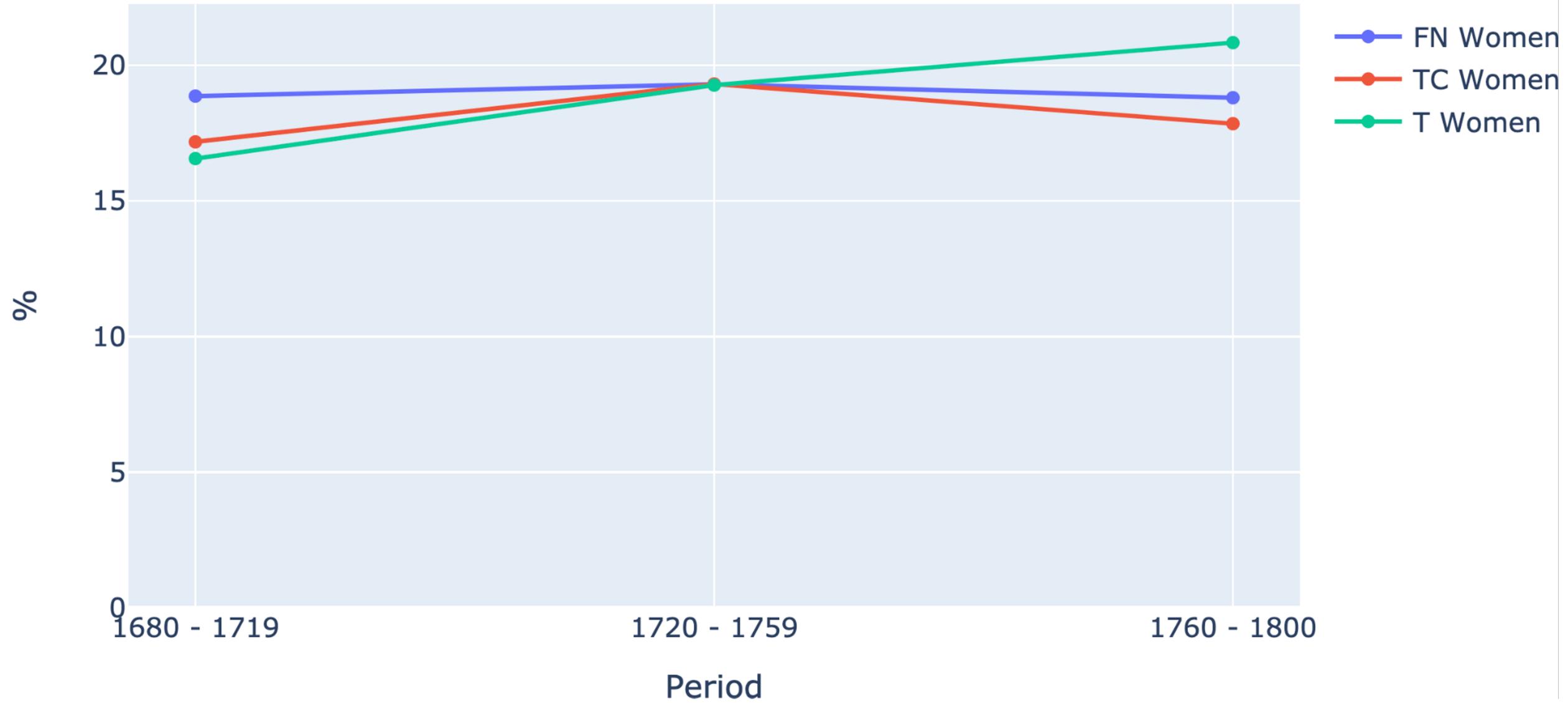
Nouns



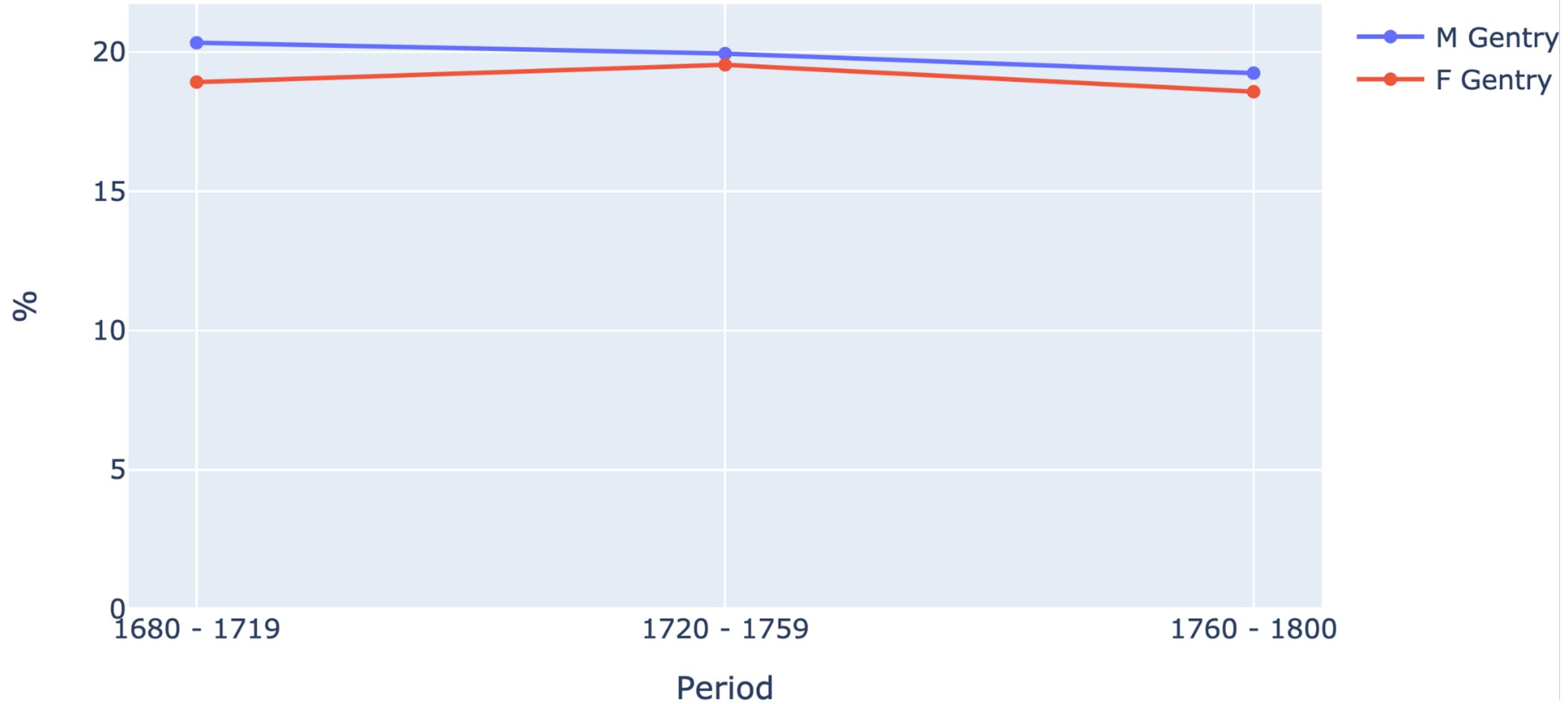
Nouns



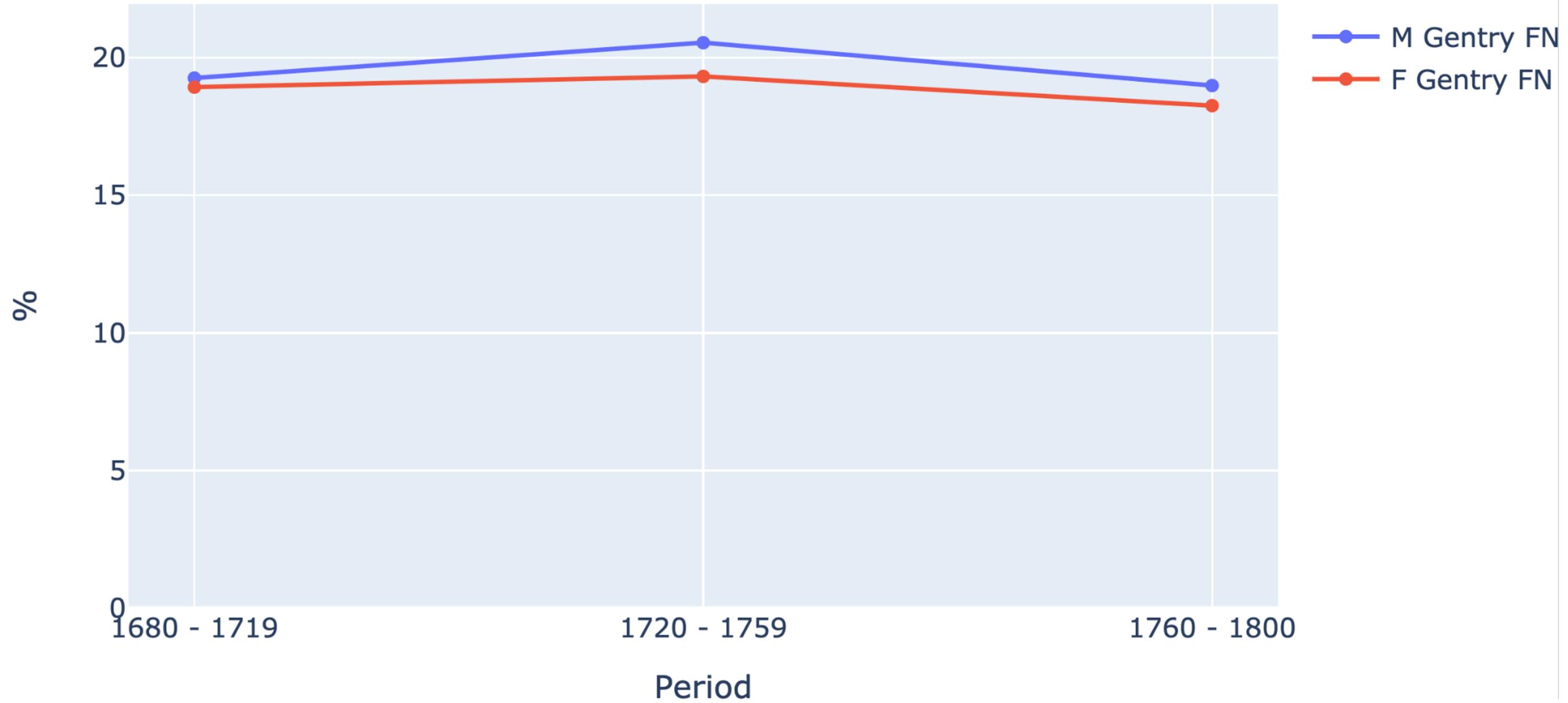
Nouns



Nouns



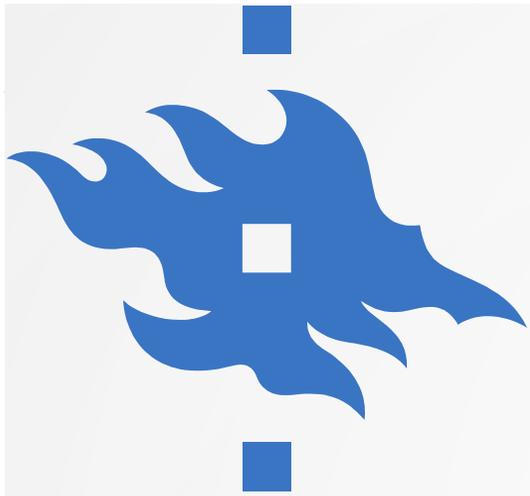
Nouns





NOUNS: SUMMARY

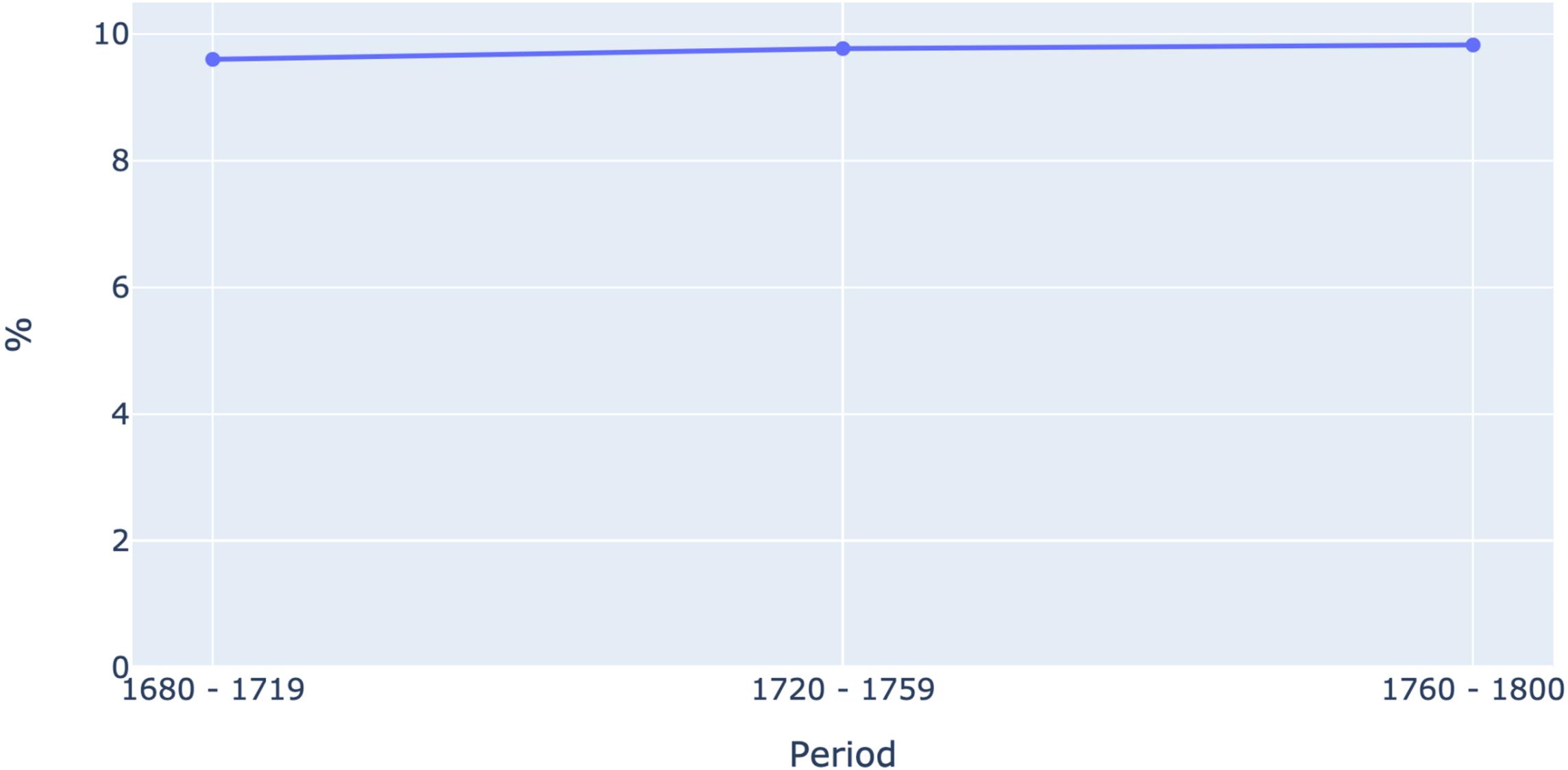
- Slight **decrease** in overall ratio
- Pretty consistent **gender** difference
 - Men > women
- **Relationship** matters!
 - Other acquaintances > nuclear family > close friends
- Middle period (1720–1759): differences sometimes disappear / slightly reversed
 - Men+women: close friends
 - Women: all relationships



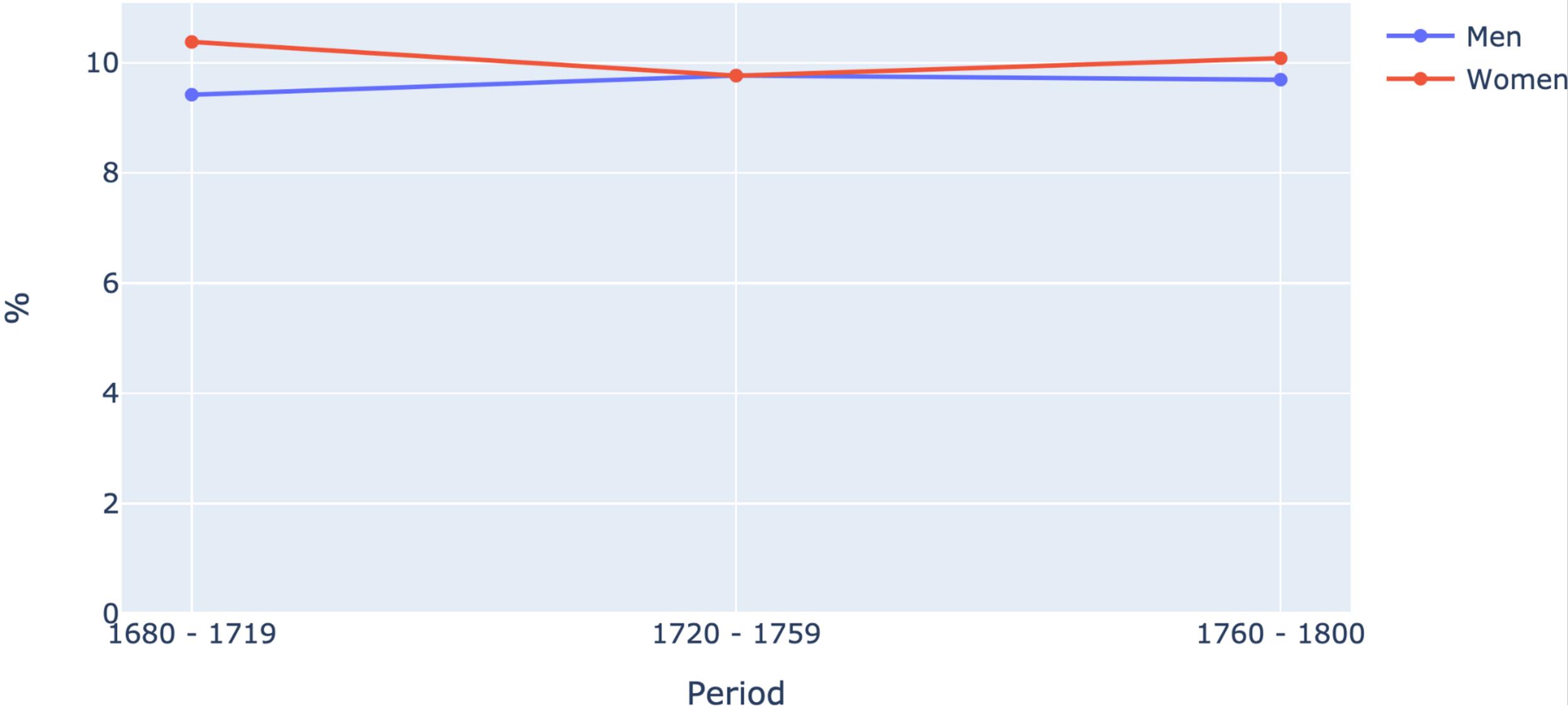
LEXICAL VERBS

+ GENDER + RELATIONSHIP + SOCIAL RANK

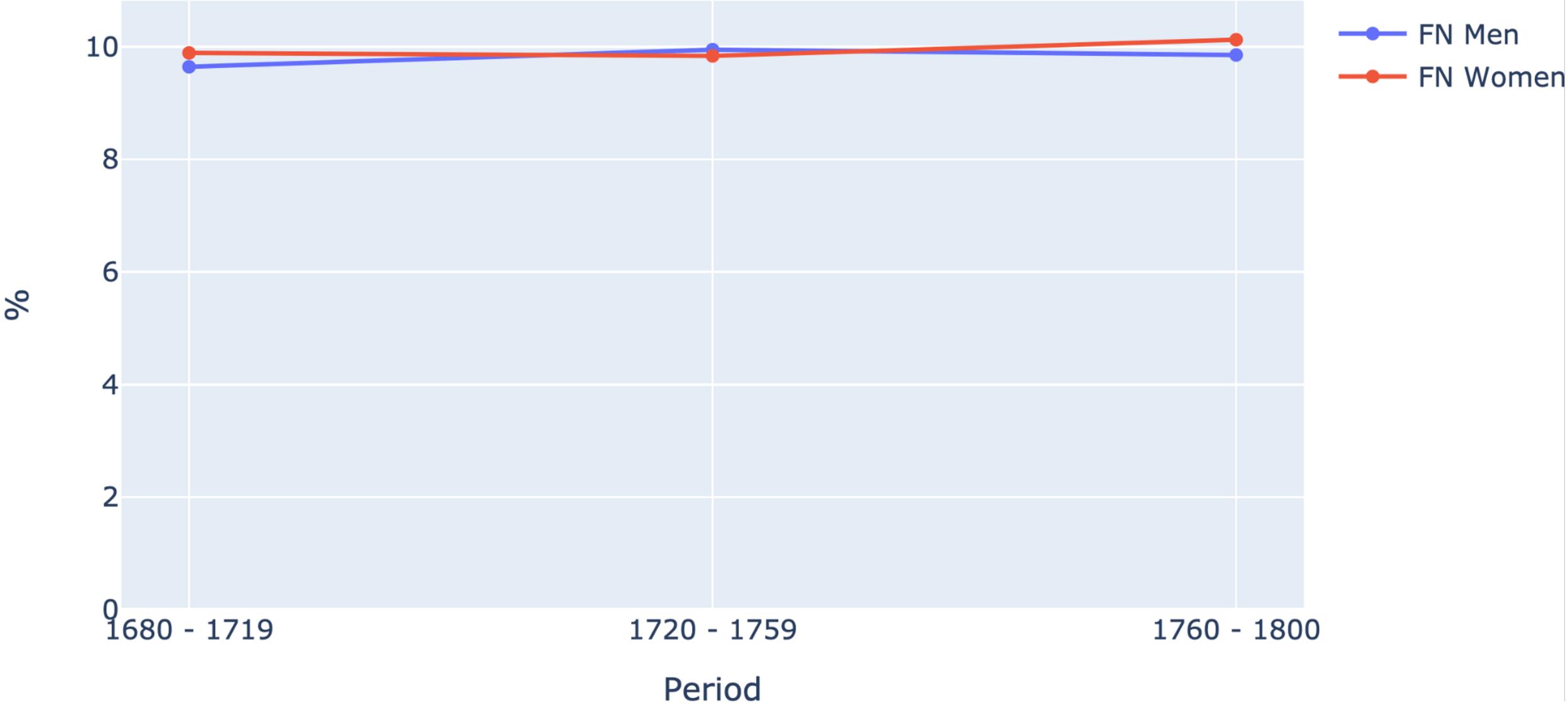
Lexical verbs (past tense excluded)



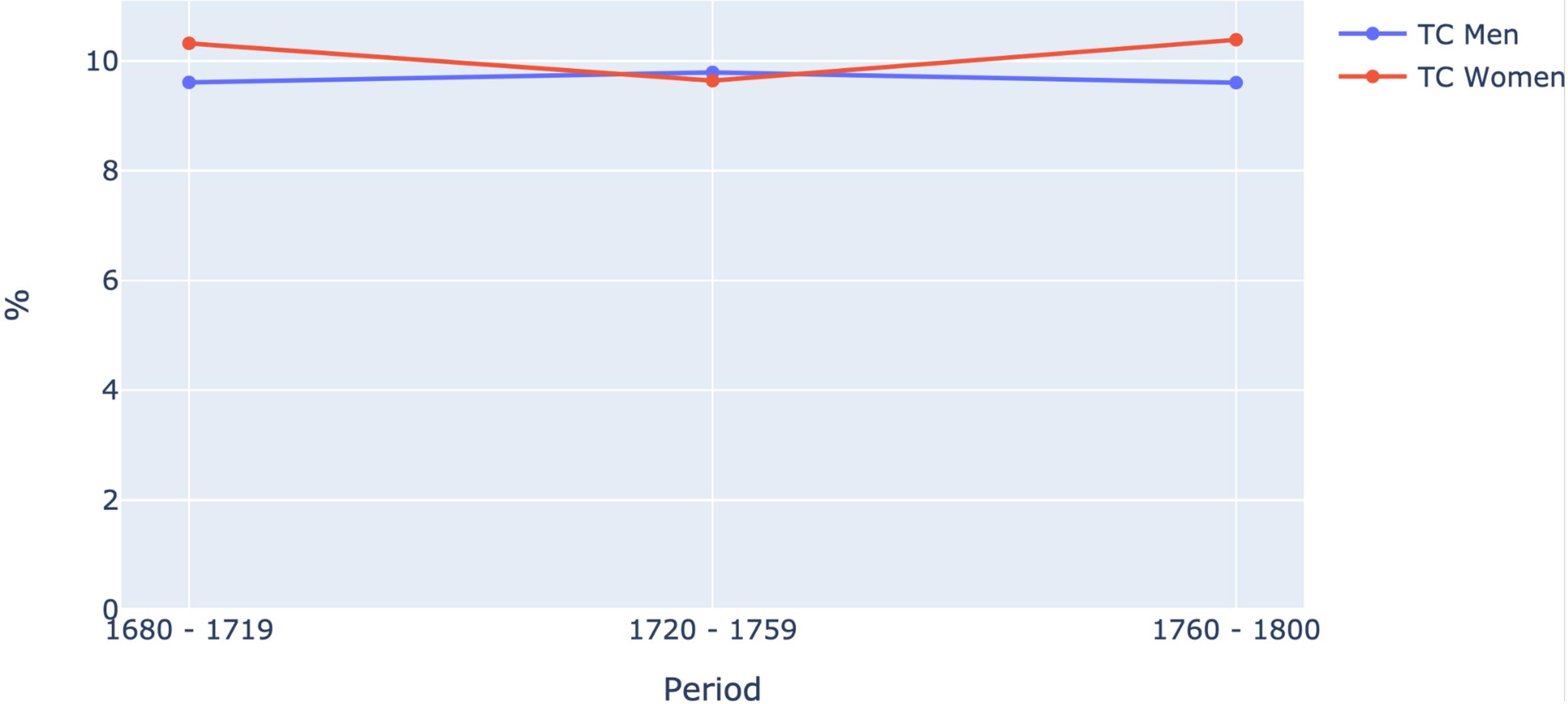
Lexical verbs (past tense excluded)



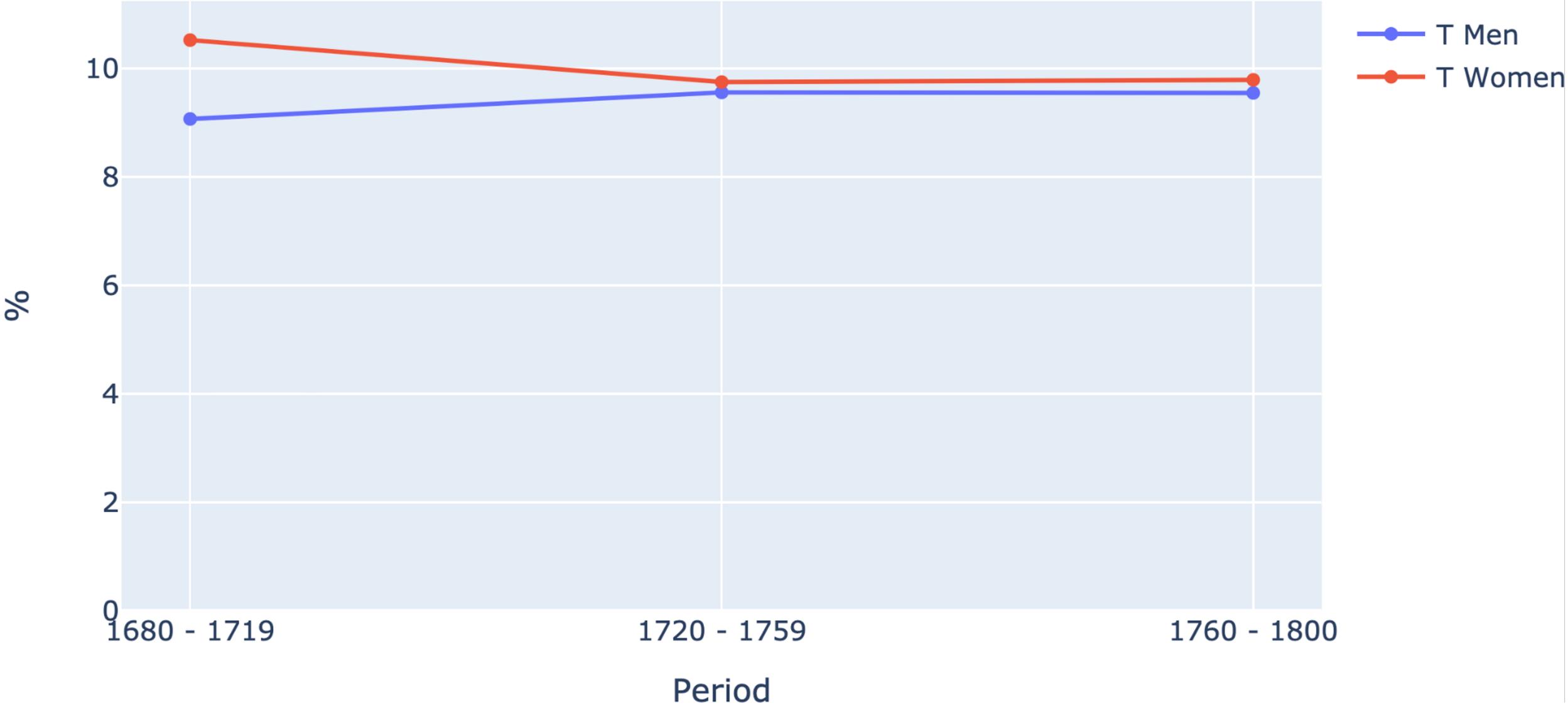
Lexical verbs (past tense excluded)



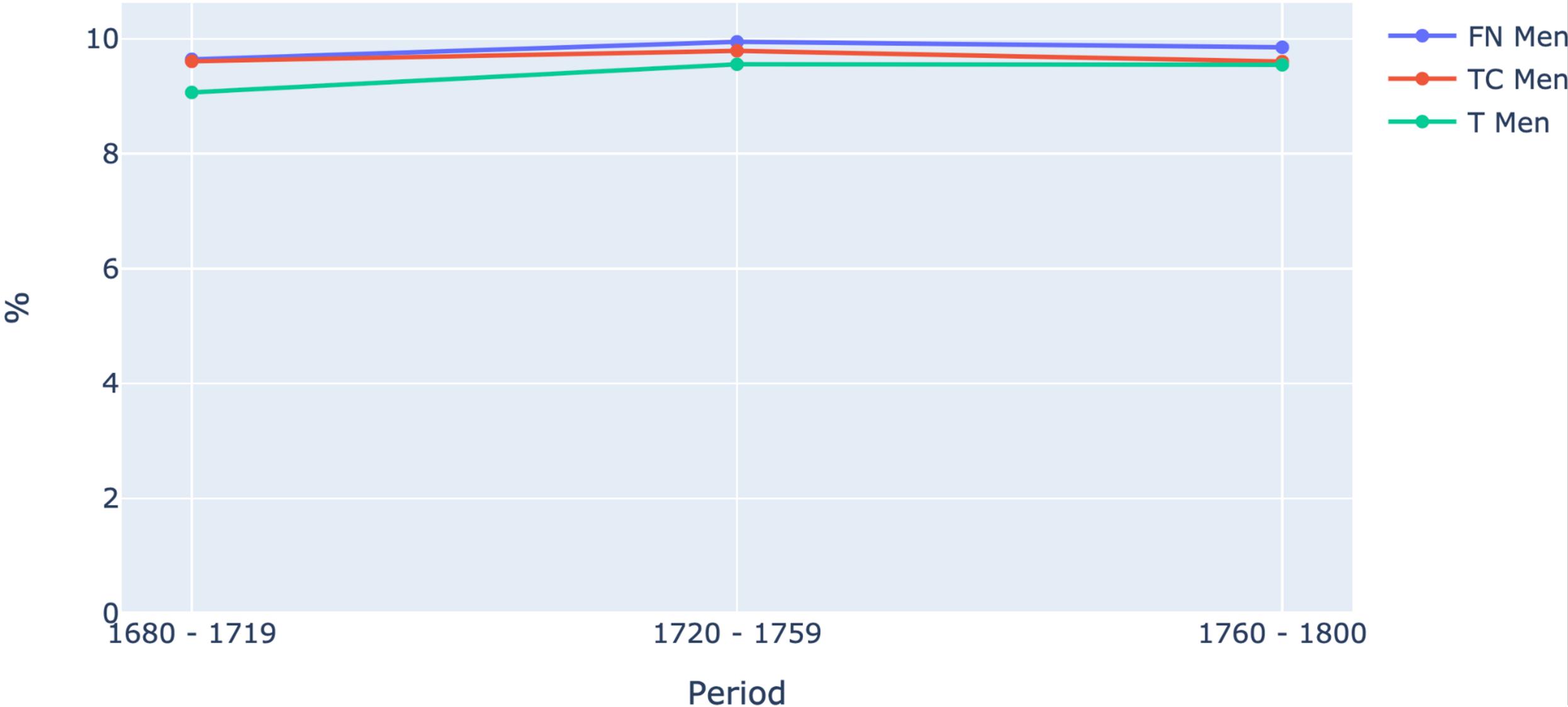
Lexical verbs (past tense excluded)



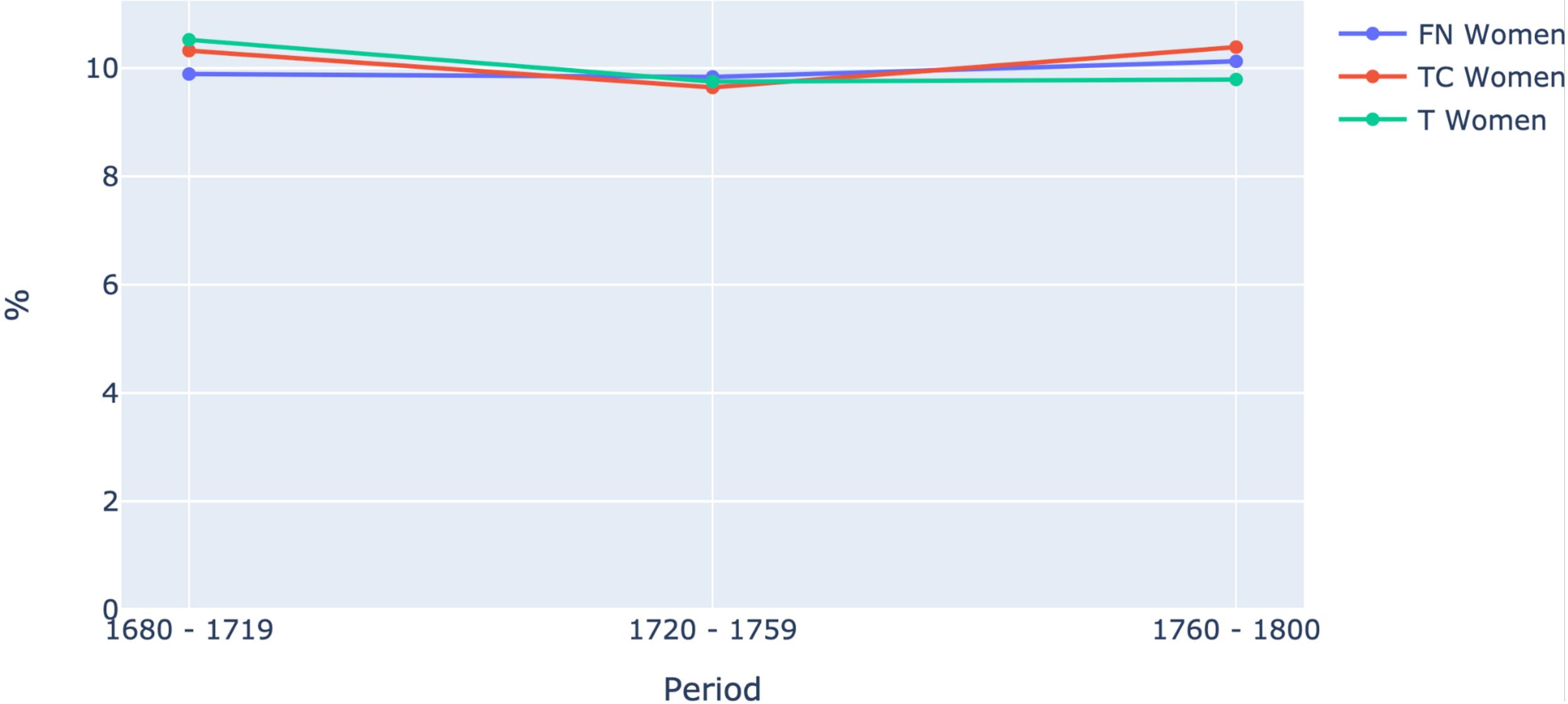
Lexical verbs (past tense excluded)



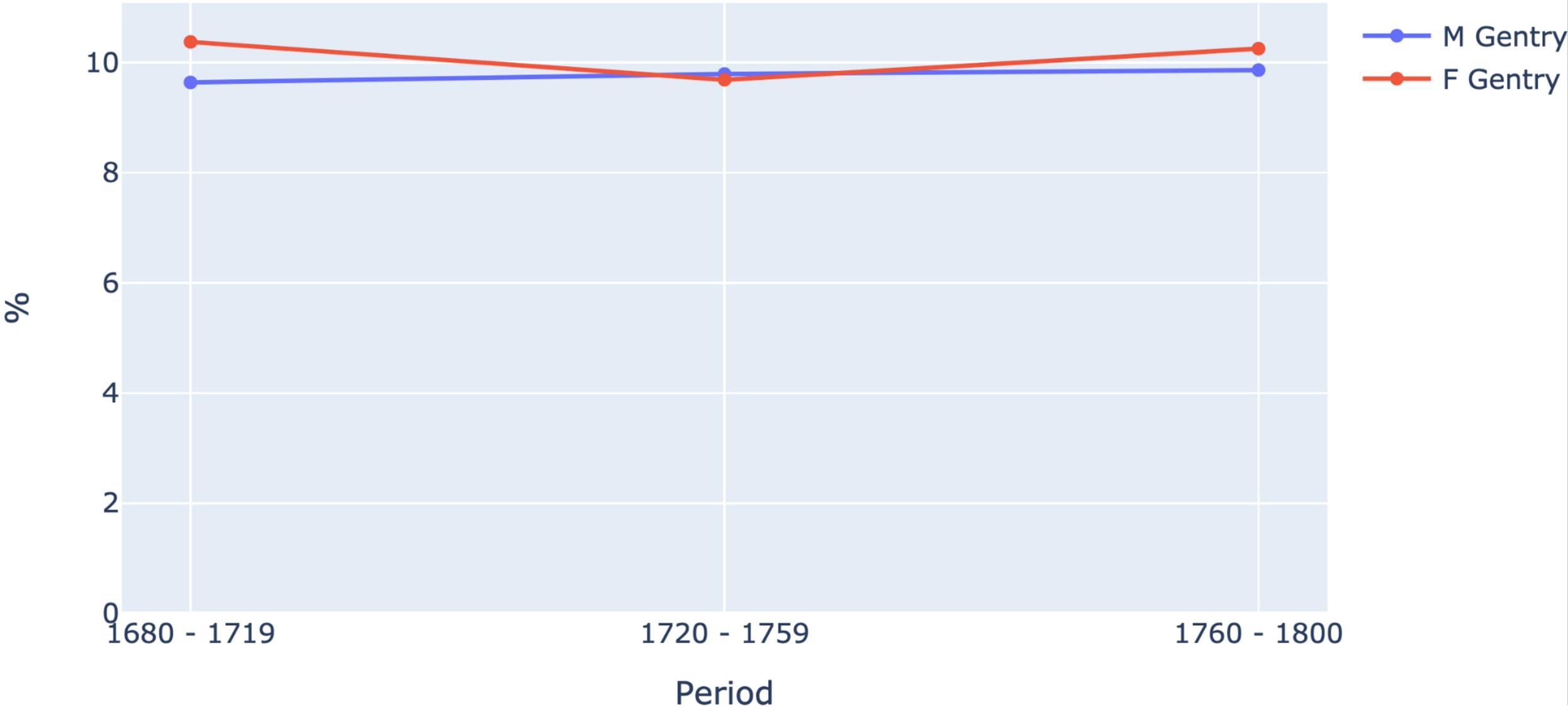
Lexical verbs (past tense excluded)



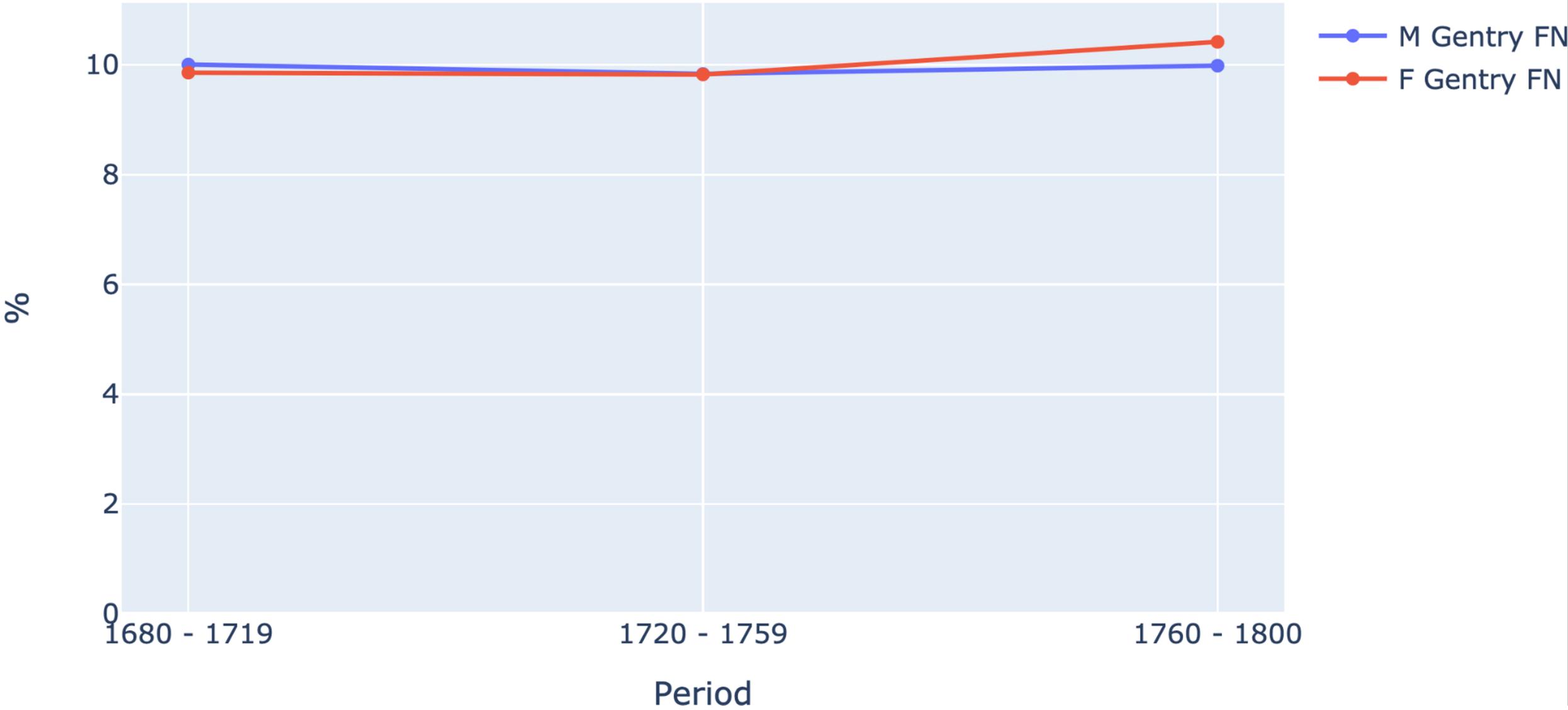
Lexical verbs (past tense excluded)



Lexical verbs (past tense excluded)



Lexical verbs (past tense excluded)





LEXICAL VERBS: SUMMARY

- Slight **increase** in overall ratio
- Somewhat consistent **gender** difference
 - Women > men
- **Relationship** matters
 - Men: nuclear family > close friends > other acquaintances
 - Women: close friends > nuclear family > other acquaintances (?)
- Middle period (1720–1759): differences often disappear / slightly reversed
 - Men+women, also with close friends + nuclear family + gentry
 - Women: all relationships (ish)



CONCLUSIONS

- **Colloquialization** of the letter genre slowly continues?
- **Gender** matters
 - On the whole, women still more involved, men more informational (but differences small)
- **Relationship** matters!
 - Close = more involved, distant = more informational
 - Also intra-gender differences (cf. Vartiainen et al. 2013)
- **Shared style**: maybe in the middle period? Role of close friends?
- More research + verification needed!



REFERENCES

- Bell, A. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.
- Biber, D. 2001. Dimensions of variation among 18th-century speech-based and written registers. H.-J. Diller & M. Görlach (eds.), *Towards a history of English as a history of genres*, 89–109. Heidelberg: Winter.
- Biber, D. & E. Finegan. 1997. Diachronic relations among speech-based and written registers in English. T. Nevalainen & L. Kahlas-Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 253–275. Helsinki: Société Néophilologique.
- Culpeper, J. & M. Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: CUP.
- Hardie, A. 2007. Part-of-speech ratios in English corpora. *International Journal of Corpus Linguistics* 12(1). 55–81.
- McIntosh, C. 2008. British English in the long eighteenth century (1660–1830). H. Momma & M. Matto (eds.), *A companion to the history of the English language*, 228–234. Chichester: Wiley-Blackwell.
- Saario, Lassi & Tanja Säily. 2020. *POS tagging the CEECE. A manual to accompany the Tagged Corpus of Early English Correspondence Extension (TCEECE)*. Helsinki: VARIENG. https://varieng.helsinki.fi/CoRD/corpora/CEEC/tceece_doc.html.
- Säily, T. 2016. Sociolinguistic variation in morphological productivity in eighteenth-century English. *Corpus Linguistics and Linguistic Theory* 12(1). 129–151.
- Säily, T. 2018. Change or variation? Productivity of the suffixes *-ness* and *-ity*. T. Nevalainen, M. Palander-Collin & T. Säily (eds.), *Patterns of change in 18th-century English: A sociolinguistic approach*, 197–218. Amsterdam: Benjamins.
- Säily, T., T. Nevalainen & H. Siirtola. 2011. Variation in noun and pronoun frequencies in a sociohistorical corpus of English. *Literary and Linguistic Computing* 26(2). 167–188.
- Säily, T., T. Vartiainen & H. Siirtola. 2017. Exploring part-of-speech frequencies in a sociohistorical corpus of English. T. Säily, A. Nurmi, M. Palander-Collin & A. Auer (eds.), *Exploring future paths for historical sociolinguistics*, 23–52. Amsterdam: Benjamins.
- TCEECE = *Tagged Corpus of Early English Correspondence Extension*. Annotated by L. Saario & T. Säily. Spelling standardized by M. Hakala et al. Compiled by T. Nevalainen et al. at the Department of Modern Languages, University of Helsinki. <https://varieng.helsinki.fi/CoRD/corpora/CEEC/>.
- Vartiainen, T., T. Säily & M. Hakala. 2013. Variation in pronoun frequencies in early English letters: Gender-based or relationship-based? J. Tyrkkö, O. Timofeeva & M. Salenius (eds.), *Ex philologia lux: Essays in honour of Leena Kahlas-Tarkka*, 233–255. Helsinki: Société Néophilologique.