

## The grammar of gender: Verb phrases with gendered subjects across 200 years of American English

Personal pronouns are indexical in multiple respects. As deictic expressions, they rely on context for their reference. However, beyond this referential indexicality, personal pronouns are also rich sites for social indexicality, as they often encode interpersonal relationships, status, or basic dimensions of social differentiation (Agha 2007).

This talk considers the social indexicality of the pronouns *he* and *she* in order to explore how gender is linguistically encoded and constructed. To this purpose, all clauses with a pronominal subject *he* or *she* (N=6,229,635) were extracted from the Corpus of Historical American English, a 475 million word corpus of written American English covering the period from 1820 until 2019 (Davies, 2021). The two pronouns are compared with respect to the overall frequency with which they occur as subjects as well as their distribution across combinations of voice, tense, and aspect. Moreover, the verbs that are strongly associated with either *he* or *she* are identified. Developments over time and in five different genres are considered.

Results show that masculine subjects are about five times more frequent than feminine ones until 1950. From this point on, a marked development towards parity can be observed; however, *he* in subject position remains about twice as frequent as *she*. Contrary to expectations, masculine pronouns show a higher percentage of passive voice subjects than feminine ones. The most strongly associated verbs for each pronoun show a clear pattern, with public (*appoint*) and cognitive (*estimate*) verbs favoring masculine subjects and verbs related to emotions (*sob*) and home life (*sew*) favoring feminine ones. However, there are important patterns of change over time in the extent to which semantic classes of verbs are associated with subject *he* or *she* (see Fig 1).

The study shows how linguistic analysis can provide insight into the cultural construction of gender and, specifically, how the incorporation of syntactic structure and verbal semantics adds value beyond simple frequency relationships (Twenge et al. 2012; Michel et al. 2011). Despite its focus on recent history, the results are relevant for current debates around pronoun use.

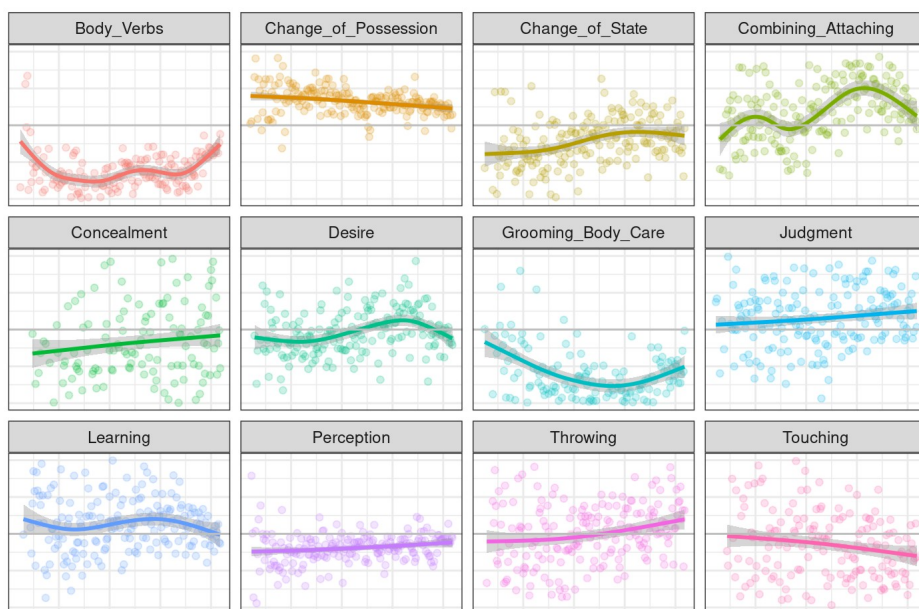


Figure 1: Gender association of different verb classes across 200 years of American English (GAM smoothers).

## References

- Agha, Asif. 2007. *Language and Social Relations*. Cambridge: Cambridge University Press.
- Davies, Mark. 2021. Corpus of Historical American English (COHA), 2021 Update. <https://www.english-corpora.org/coha/>.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Google Books Team; Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176–182.
- Twenge, Jean M., W. Keith Campbell & Brittany Gentile. 2012. Male and female pronoun use in U.S. books reflects women's status, 1900–2008. *Sex Roles* 67: 488–493.