

Gender representation in linguistics journals

Sarah Babinski, Rikker Dockum, Chris Geissler, Hadas Kotek

Linguistics exhibits pervasive gender bias and tends toward stereotypes in example sentences, a major source of data in the field.

Linguistic examples as data

Linguists rely on **example sentences** as a main source of data.

- (1) **English word order is SVO:** (2) **Japanese is SOV:**
- John read a book. John-wa hon-o yon-da
 - * John a book read. John book read

➤ **Example sentences contain implicit biases** (Macaulay and Brice, 1994, 1997; Pabst, Cepeda, Kotek, and Syrett, 2019).

Bias in linguistics textbooks

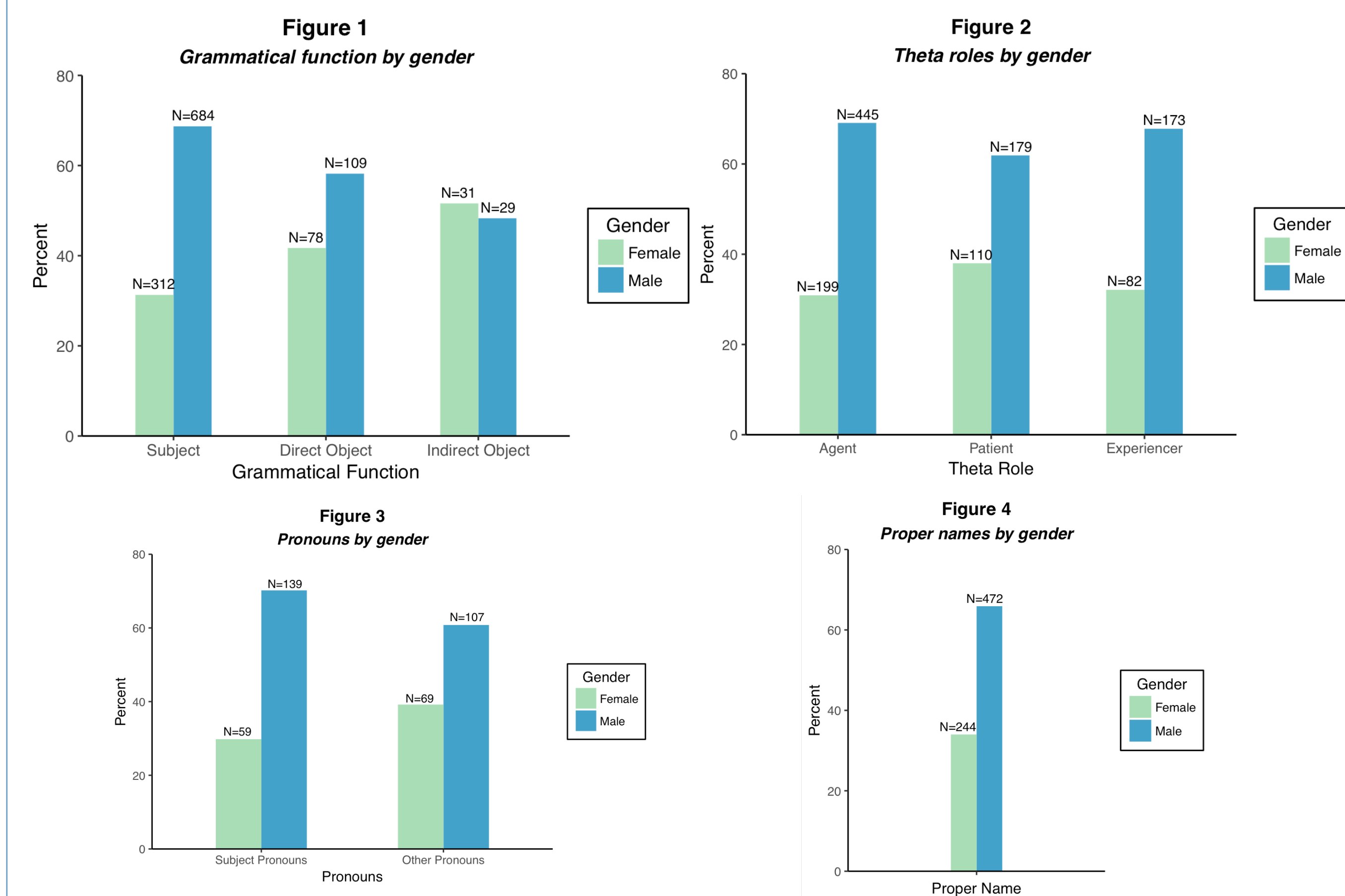
Macaulay and Brice (1997): An analysis of eleven syntax textbooks.

“The **majority** of constructed example sentences in syntax textbooks are **biased toward male-gendered noun phrases**, and ...contain **highly stereotyped representations** of both genders.”

Pabst et al. (2019): Replicates this study with six current textbooks

200 examples were sampled from each textbook. Nominals coded for:

- ✦ **Grammatical gender** (female, male, other)
- ✦ **Grammatical function** (subject, direct object, indirect object, etc.)
- ✦ **Thematic role** (agent, patient, experiencer, recipient, etc.)
- ✦ **Lexical choices** (pronouns, proper names, violence, appearance,...)



Men...

- ✦ Appear more often as **arguments** in general
- ✦ Are more likely to be **subjects** and **agents**
- ✦ Have pronouns and **proper names** mentioned more often
- ✦ Engage in **'intellectual activities'** (e.g. book reading) more often
- ✦ Have **occupations** more often; have a wider range of occupation
- ✦ Perpetrate **violence** more often and more severely than women

Bias in linguistics journals?

Research question: Do linguistics journals exhibit gender bias in example sentences in the same ways as in linguistics textbooks?

Benefits of this approach:

- ✦ Journals are available electronically
- ✦ Articles cover a wide timespan
- ✦ The readership for journals is bigger (and a different audience)
- ✦ Much more data, leading to robust conclusions

Data and methods

Our corpus:

- 4 linguistics journals (data from 2 presented here)
- Scraped and extracted text from all articles 1997-present
- Example sentences automatically extracted from text

Yale undergraduate linguistics students have analyzed:

- 4290 example sentences, yielding:
- 582 instances of female arguments
- 1453 instances of male arguments

Stereotypes in example sentences

“Her heart desires especially to die together with him.”

“The girl can cook fish very well. (despite expectations)”

“A dentist’s wife entered the room. He (the dentist) wasn’t there.”

“the fact that John is good at math”

“How do you want Pedro to fix the car?”

“She saw him kissing all the girls”

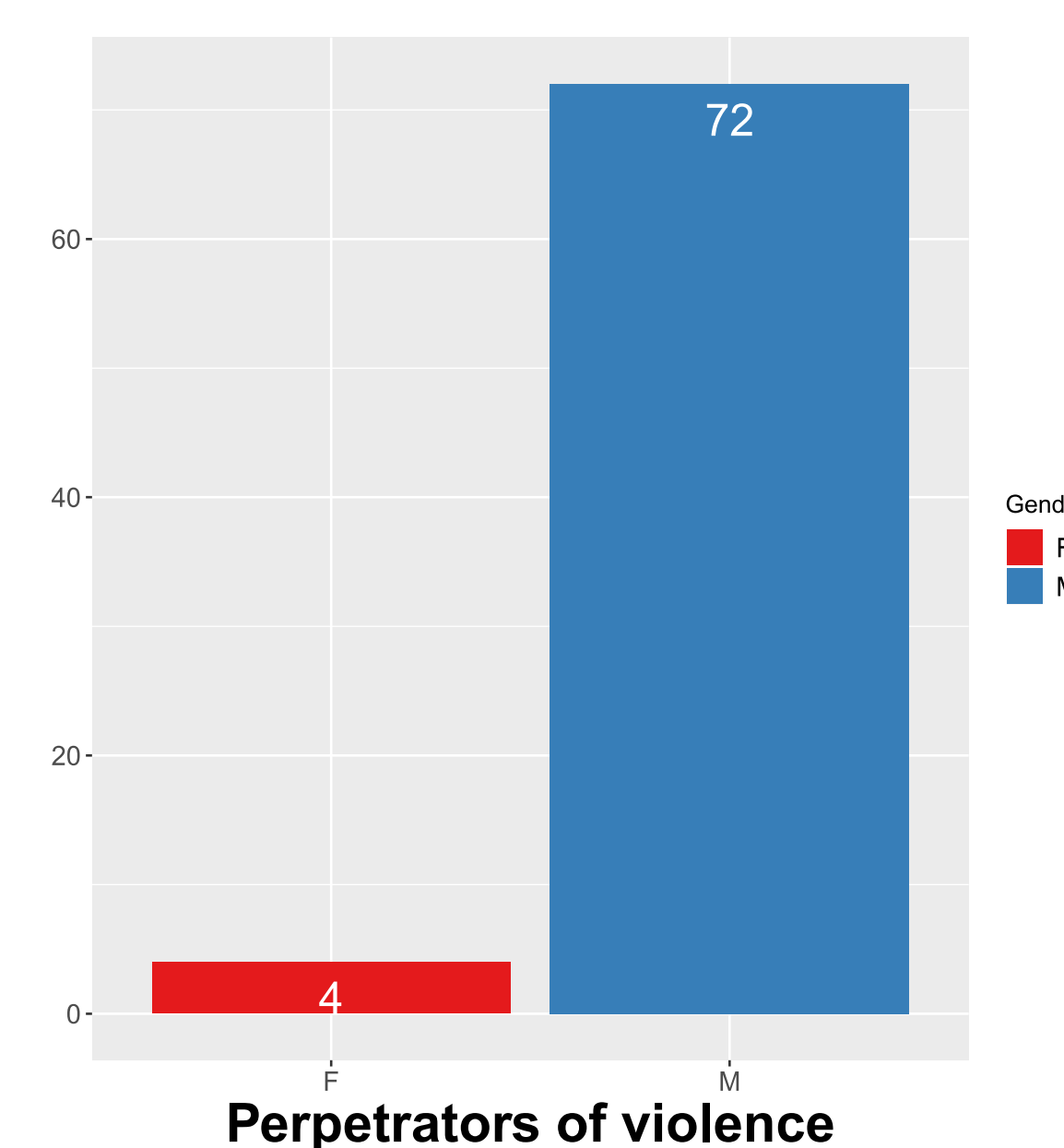
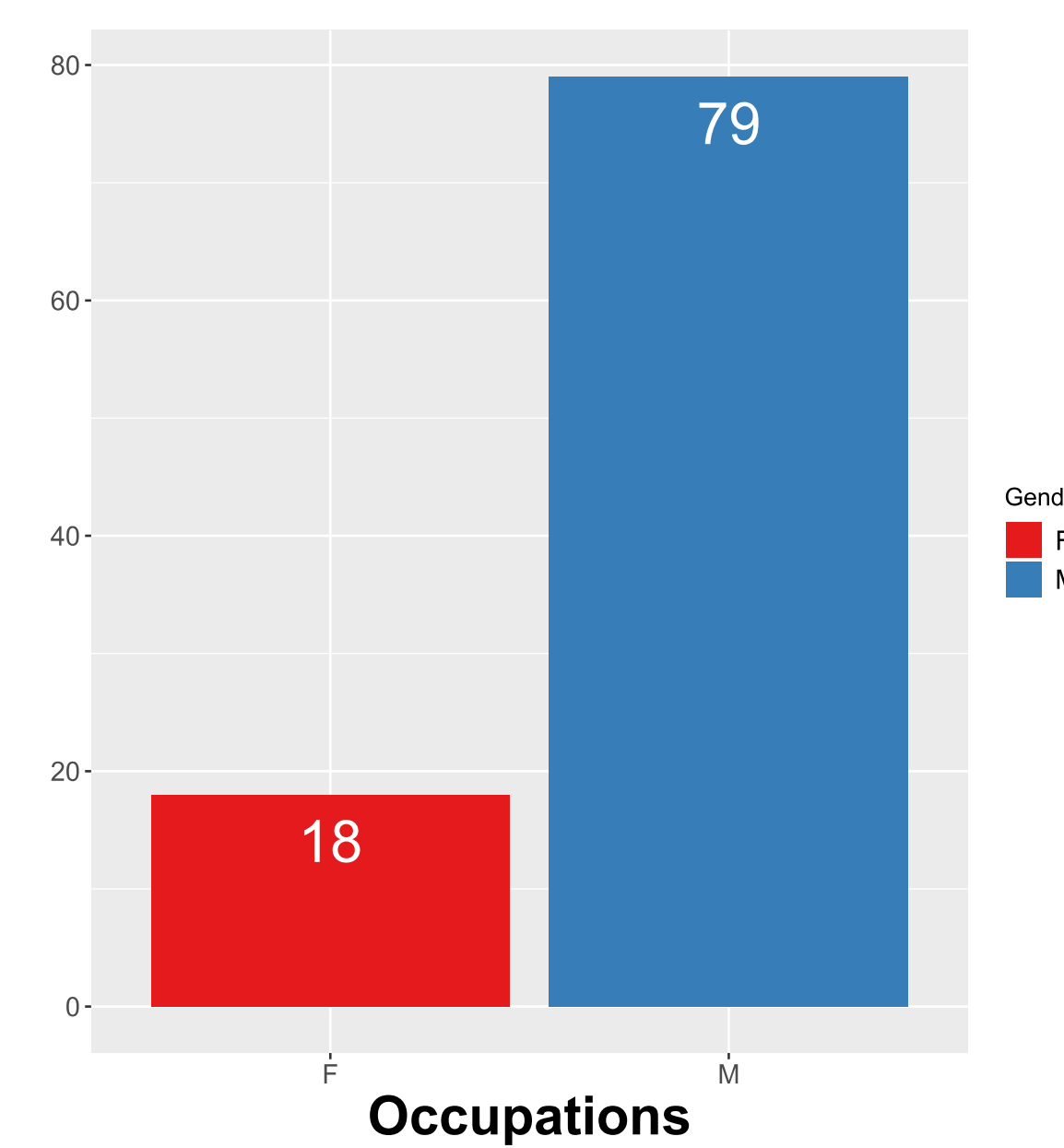
“Bill seems to John to be a genius.”

“Hari said that he hit Rashmi”

“Divakara lost his mind because of Divya’s beauty.”

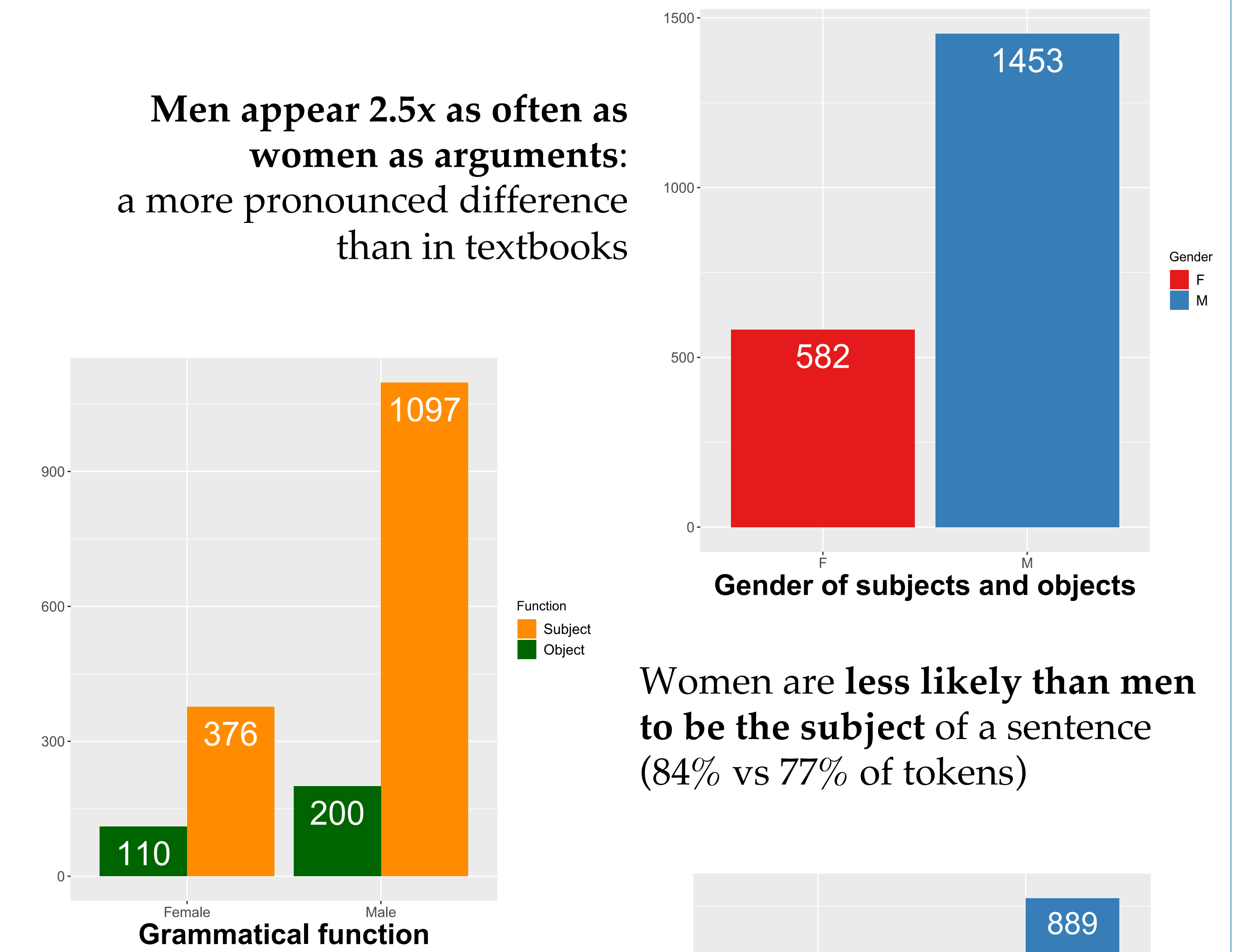
“It’s amazing how many cars he owns.”

“John wants Mary to eat either rice or beans.”



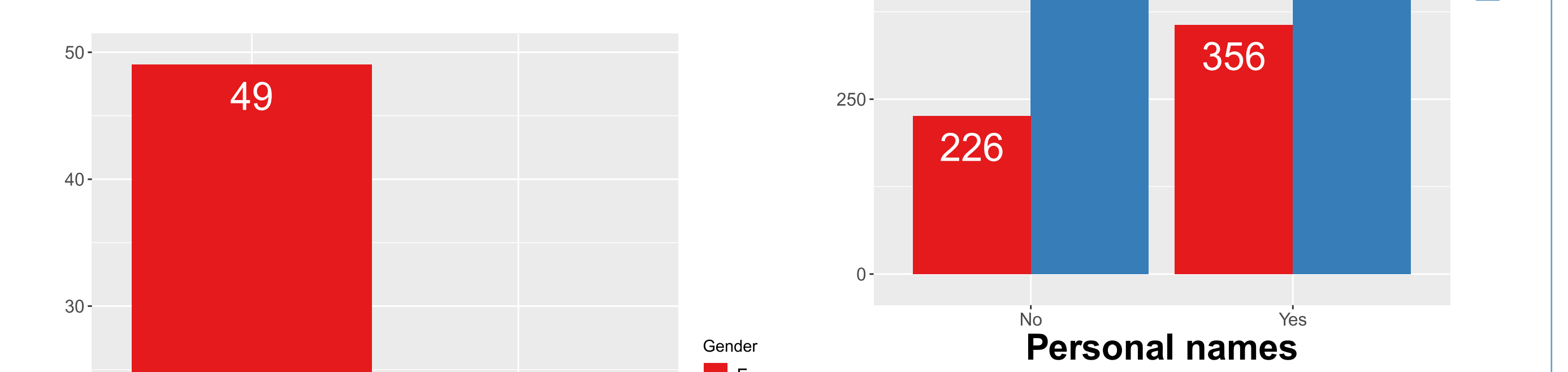
Preliminary results

Men appear 2.5x as often as women as arguments: a more pronounced difference than in textbooks

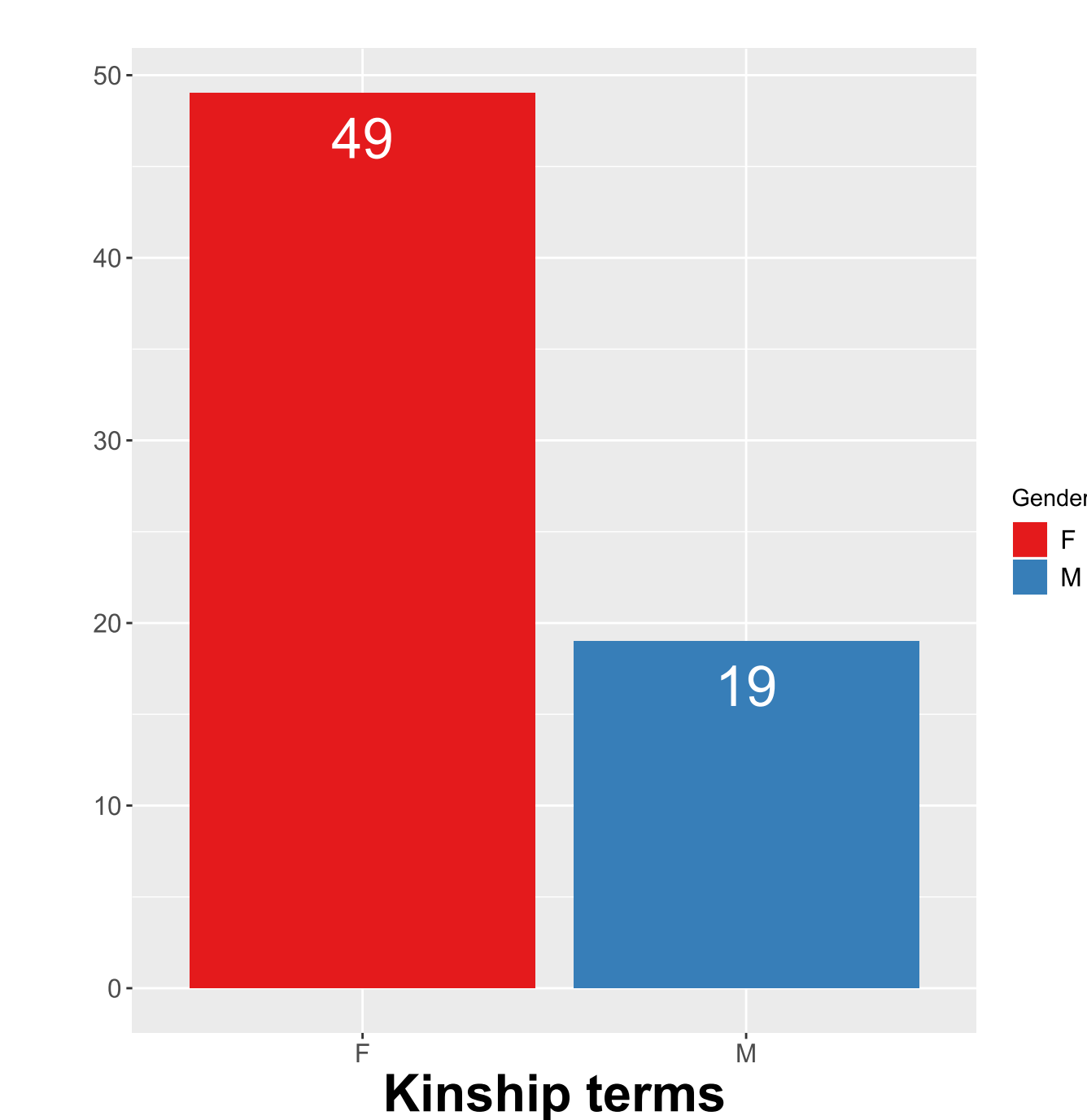


Women are less likely than men to be the subject of a sentence (84% vs 77% of tokens)

Women and men referred to by proper names at the same rate (61%)



Women are 6.5x more likely to be referred to by a kinship term (mother, sister, etc.) when scaled for corpus frequency



Next steps

- ✦ Statistical analysis
- ✦ Further possible factors of interest:
 - Example type? (elicited vs. corpus)
 - Gender of author?
 - Year (or decade) of journal?
 - Language of example sentence?