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:

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

Bias in linguistics 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, ...)

Data and methods

Our corpus:

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

Yale undergraduate linguistics students have analyzed:

- 4290 example sentences, yielding:
  - 1453 instances of male arguments
  - 582 instances of female arguments
  - 376 instances of female arguments from 2 presented journals

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?