Gender representation in linguistic example sentences

Hadas Kotek, Rikker Dockum, Sarah Babinski,
Christopher Geissler

Linguistic Society of America
January 2020

Introduction

In 1996, the LSA publishes the *Guidelines for Non-Sexist Usage*.

Macaulay and Brice 1997, *Language*: An analysis of eleven syntax textbooks published 1969–1994.

> *"The **majority** of constructed example sentences in syntax textbooks are **biased toward male-gendered NPs**,and ...contain **highly stereotyped representations** of both genders."*

20 years later, Pabst, Cepeda, Kotek, and Syrett (LSA, 2018) report similar results for a study of six syntax textbooks published 2005-2017.

Introduction

Keep in mind...

- People who identify outside of the gender binary may or may not adopt gendered language to refer to themselves.
- This has nothing to do with the sex they were assigned at birth.

Introduction

Today: A study of gender representation in journal papers published between 1997–2018 in *Language*, *Linguistic Inquiry*, and *Natural Language & Linguistic Theory*.

- Do the biases found in syntax textbooks extend beyond this limited genre and into scholarly work in linguistics?
- ...and what can we do about it?

Roadmap

§1 Introduction

§2 Background: Macaulay and Brice (1997)

§3 Gender representation in journal papers

§4 Discussion: Why does this matter? How can we improve?

Roadmap

§1 Introduction

§2 **Background: Macaulay and Brice (1997)**

§3 Gender representation in journal papers

§4 Discussion: Why does this matter? How can we improve?

Macaulay & Brice (1997): Overview

Comparative study of constructed examples from 11 syntax textbooks published between 1969 and 1994.

- **Study 1**: 1,032 examples from one textbook ('Syntax textbook,' 1991); male author (published inMacaulay and Brice 1994)
- **Study 2**: 10 additional textbooks published between 1969 and 1994 to generalize results. 7 male authors, 3 female authors.

Macaulay & Brice (1997): Methods

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

- **Grammatical gender** (female, male, other)
- **Grammatical function** (subject, DO, IO, etc.)
- **Theta roles** (agent, patient, experiencer, recipient, etc.)
- **Lexical choices** (pronouns, proper names, violence, appearance, reading and writing, etc.)

Macaulay & Brice (1997): Results

Men...

- Appear more often as **arguments** than women
- Are more likely to be **subjects** and **agents** than women
- Are **subjects** and **agents** more often than other arguments
- Have **pronouns** mentioned more often than women
- Have **proper names** twice as often as women
- Are engaged in '**intellectual activities**' (book reading/handling) and handle **cars** more often than women
- Are described as having **occupations** more often than women, and in a wide range of occupations
- Perpetrate **violence** more often than women

Macaulay & Brice (1997): Results

Women...

- Often lack names, but are referred to with **kinship terms** (*X's wife, mother*) more often than male arguments are
- Have their **appearance** described more often than male arguments

Macaulay & Brice (1997): Results

Some differences based on **author sex**:

YEAR	AUTHOR SEX	SUBJECT			AGENT		
		3F (%)	3M (%)	Other (%)	3F (%)	3M (%)	Other (%)
1969	M	6	24	69	8	38	54
1975	M	10	41	49	9	53	38
1979	M	20	33	47	24	38	38
1984	M	11	47	42	15	51	34
1986	M	4	49	47	2	47	51
1988	M	8	36	57	6	43	51
1994	M	14	29	57	17	35	48
1992	F	41	22	37	47	20	33
1993	F	28	21	52	30	21	49
1994	F	6	39	55	6	42	52

Macaulay & Brice (1997): Selected examples

(1)
a. Every painting of Maja and photograph of Debbie pleased Ben.
b. Harry watches the fights and his wife the soap operas.
c. Bill is proud of his father and tired of his mother.
d. John might drown the kittens/his wife/??his goldfish/!his frying pan/!his birth.
e. Steven likes but Maja hates the man next door.
f. We consider him to be a genius and her to be a fool.
g. The man who shot her believed there was someone else who was seeing Helen.
h. His wife saw Hercule, her husband.
i. The man killed, cut up, and ate his children

...and so many more

Macaulay & Brice (1997): Selected examples

In addition, **explicit and suggestive language**:

(2) a. Max doesn't beat his wife because he loves her.

b. She's fond of John naked. (structural ambiguity)

c. After Rambo as a lover, she was exhausted.

d. I can't imagine you in kinky boots.

e. Personally, inflatable dolls bore me.

f. She'll soon tire of her sexploits.

g. What a nice pear Mary's got!

h. John forced Mary to be kissed by Bill.

i. He once glonked an out-of-work actress.

j. The lascivious tree who we saw in the magic forest waved his luxuriant branches lustfully at Mary and said, 'You can fondle my foliage anytime, darling'.

...and many more

Macaulay & Brice (1997): Conclusion

> *"Our results clearly illustrate the need for such scrutiny:* ***females are simply not significant actors in the world constructed by sample sentences****." (p. 816)*

We might add...

- Neither are non-binary individuals.
- Very little has changed in more recent textbooks (Pabst, Cepeda, Kotek, and Syrett, 2019).

Roadmap

§1 Introduction

§2 Background: Macaulay and Brice (1997)

§3 **Gender representation in journal papers**

§4 Discussion: Why does this matter? How can we improve?

Gender representation in journal papers

Textbooks are a very specific genre.

- Is this true of Linguistic research more generally?

We examine all papers from 3 journals: *Linguistic Inquiry, Natural Language & Linguistic Theory, Language* between the years 1997–2018.

$\Rightarrow$ 873 papers in total; 15,826 3rd person arguments

Methods

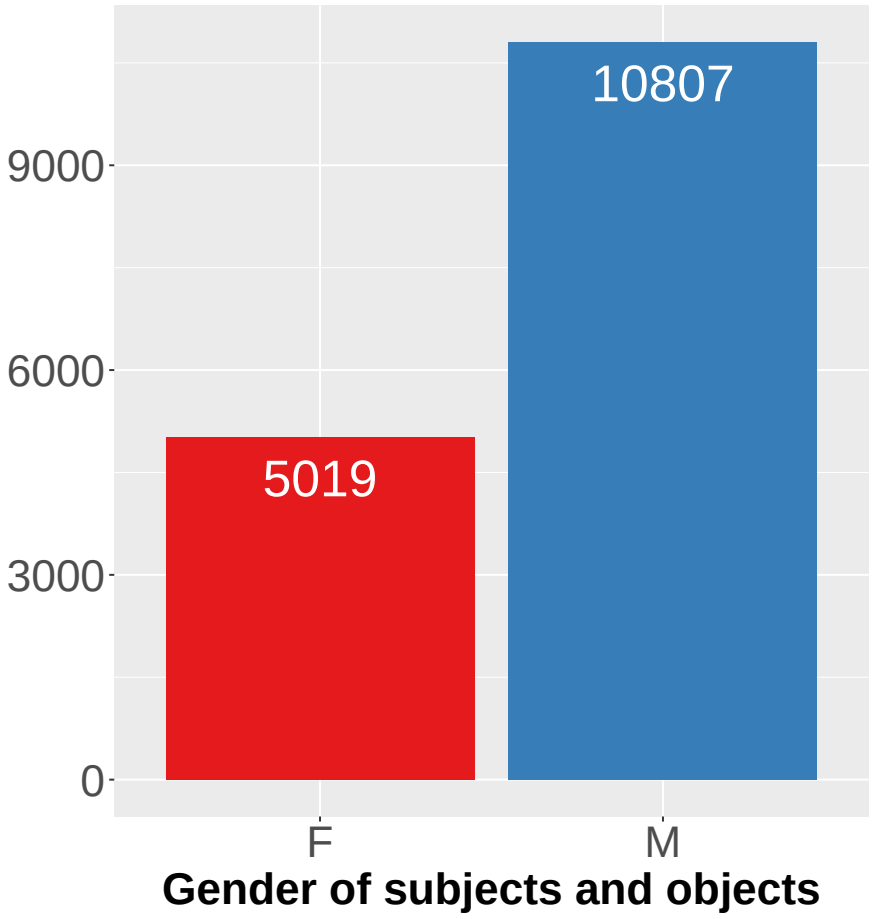
Examples extracted using Regular Expressions.

Similar coding to textbook project. 20 Yale undergrads hired to code.

Some coding relegated to automated tools:

- author gender (NLTK gender classifier),
- positive/negative emotions (sentiment analysis),
- specific tokens: kinship, violence, appearances, cars, intellect (Regular Expressions).

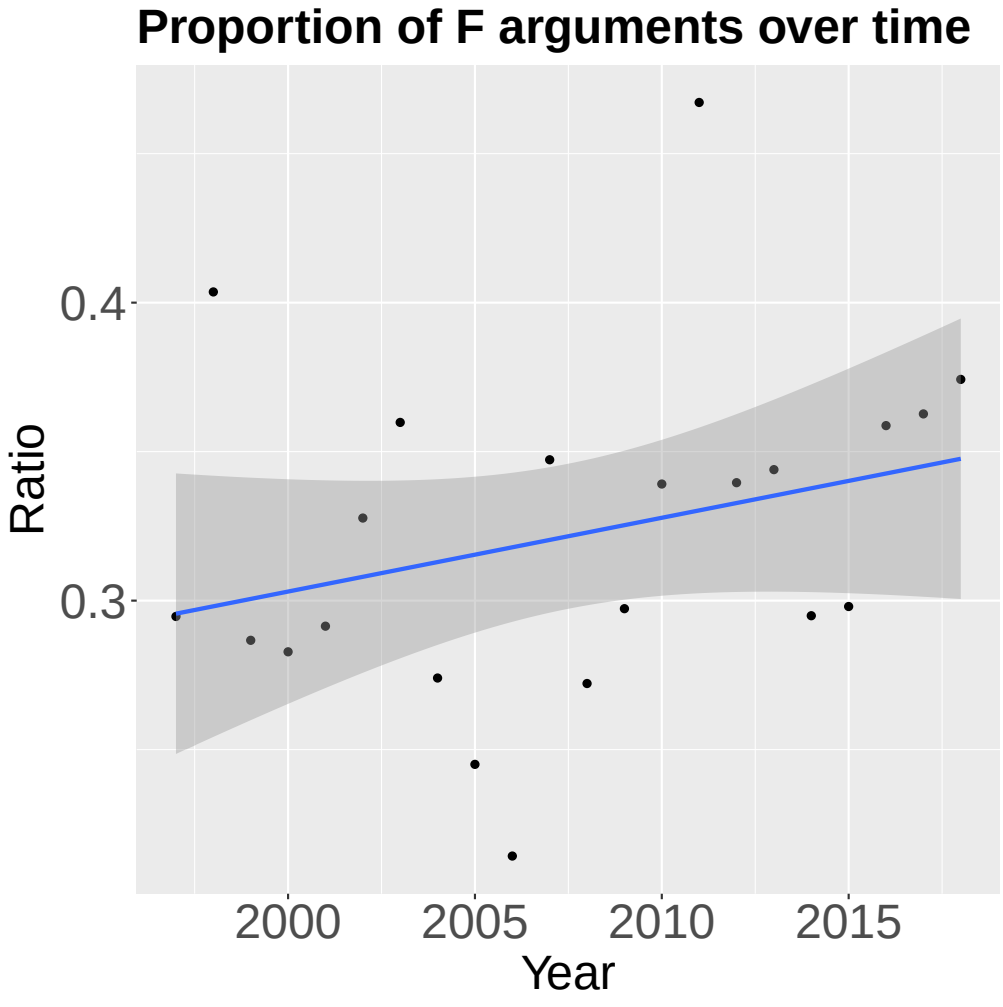
Arguments at a glance



Female-gendered NPs represent 32% of the total arguments (15,826)
(Namely, a ratio of 2.1 male arguments for every 1 female argument).

Arguments over time

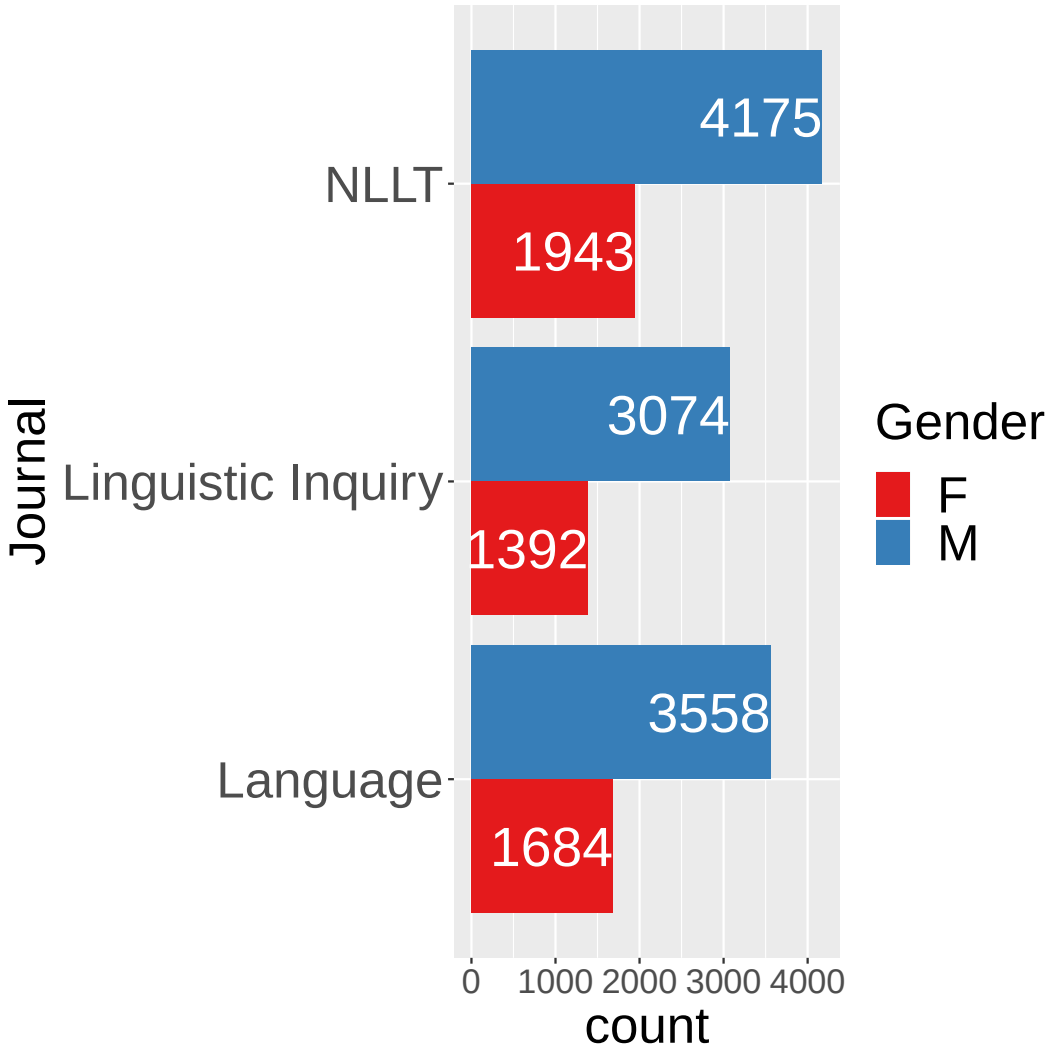
An ever so slight improvement over the 20 years we studied:



Arguments by journal

By journal: the same trends.

We'll show collapsed graphs throughout.

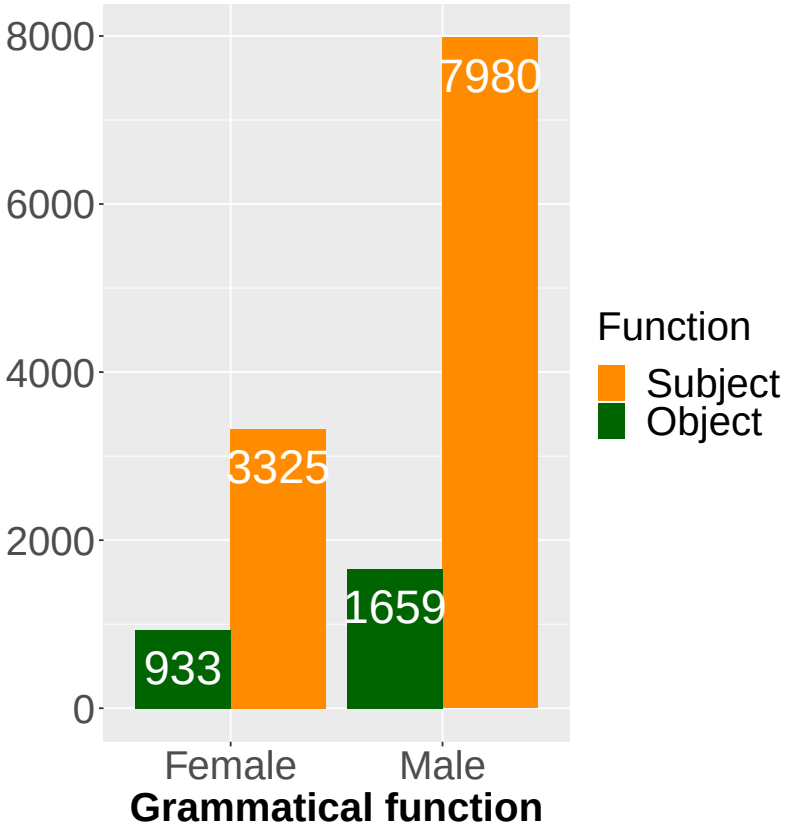


NLLT: 31%

LI: 32%

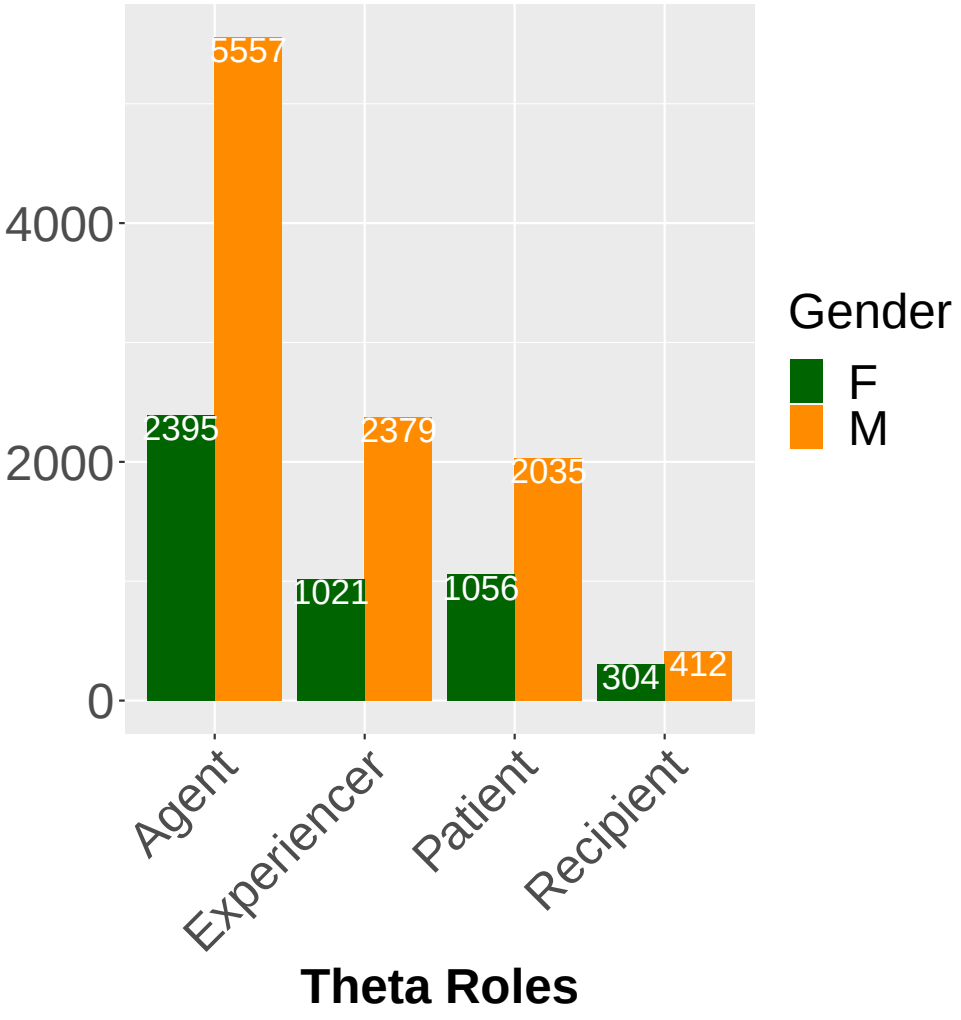
Language: 32%

Grammatical Function



74% of male arguments vs 66% of female arguments are **subjects**.

Theta roles



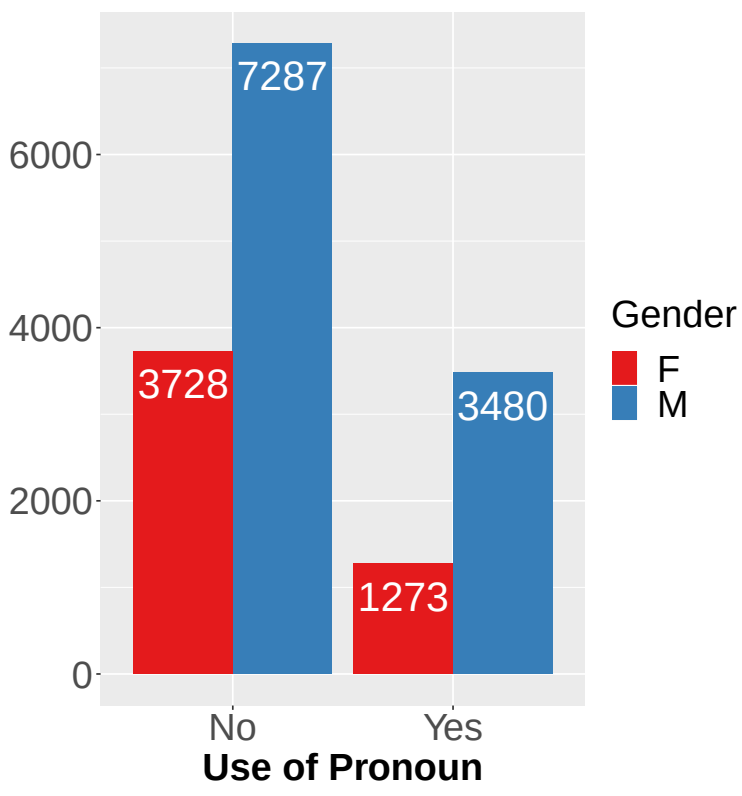
Agents: 30% female

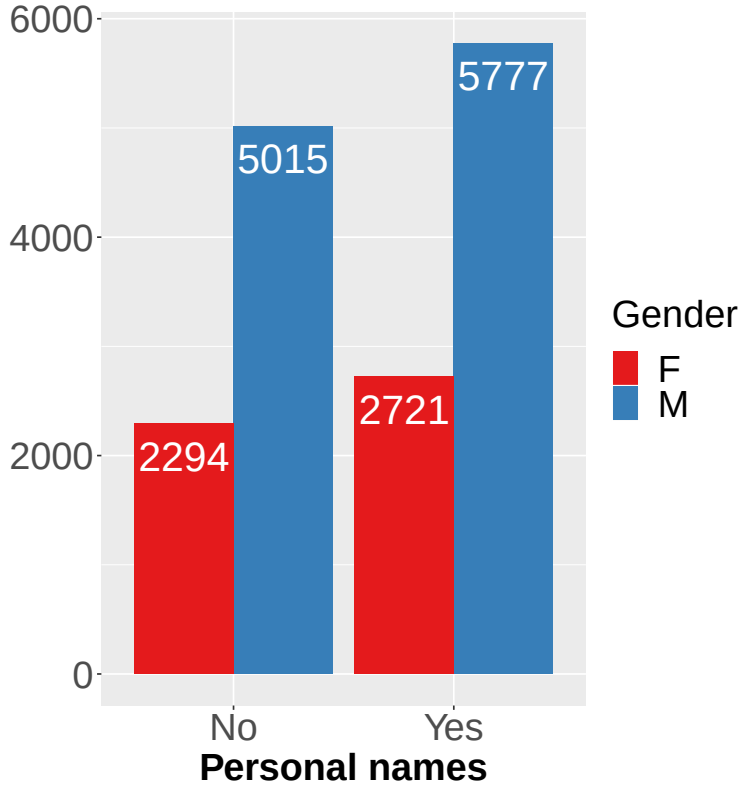
Experiencers: 30% female

Patients: 34% female

Recipients: 43% female

Pronouns and proper names

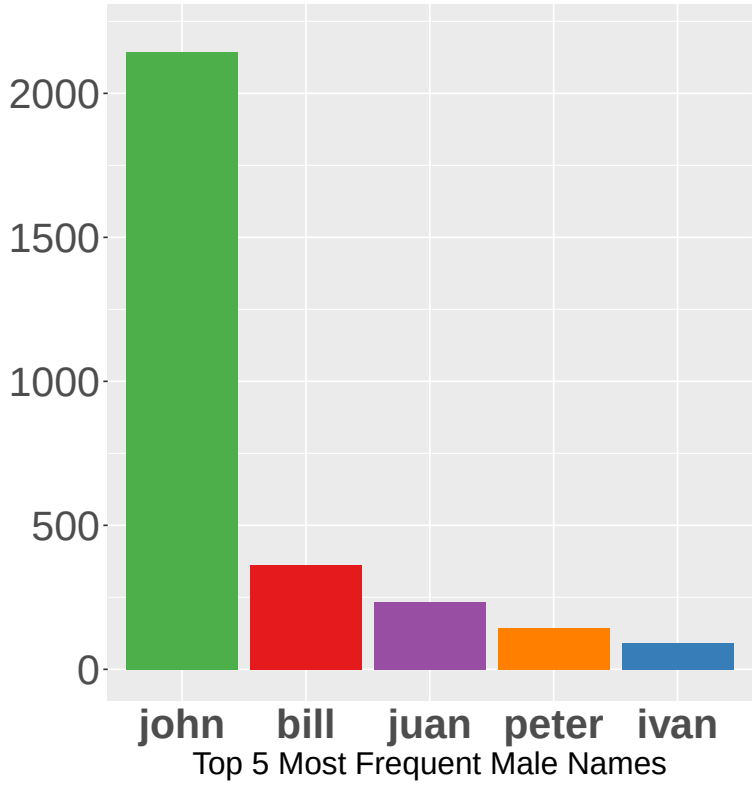




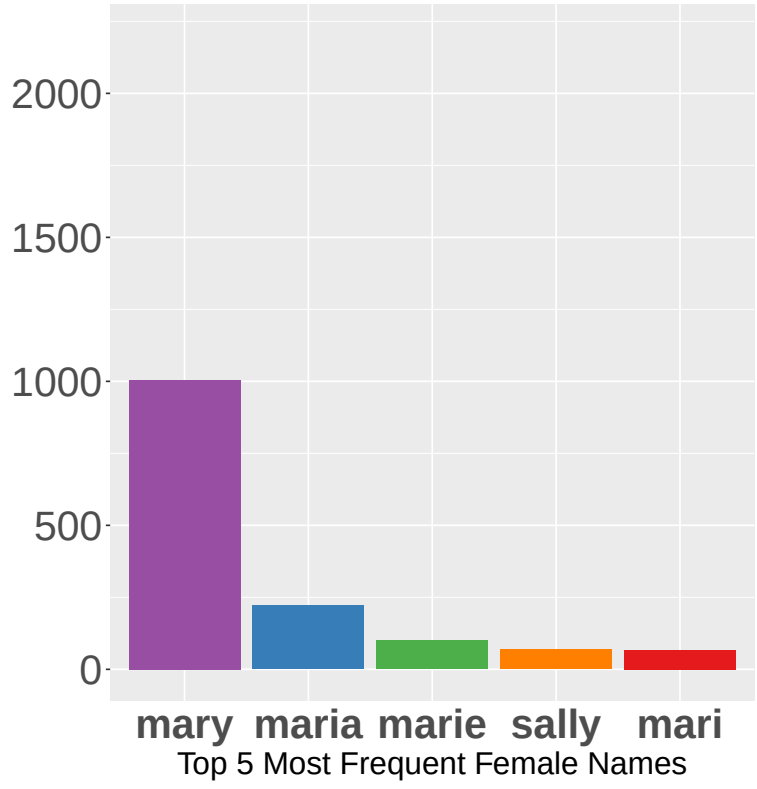
Male and female-gendered NPs have **pronouns** 32% vs 25% of the time.

They are **proper names** 53% vs 54% of the time.

Proper names: Fun facts



Top 5 Most Frequent Male Names

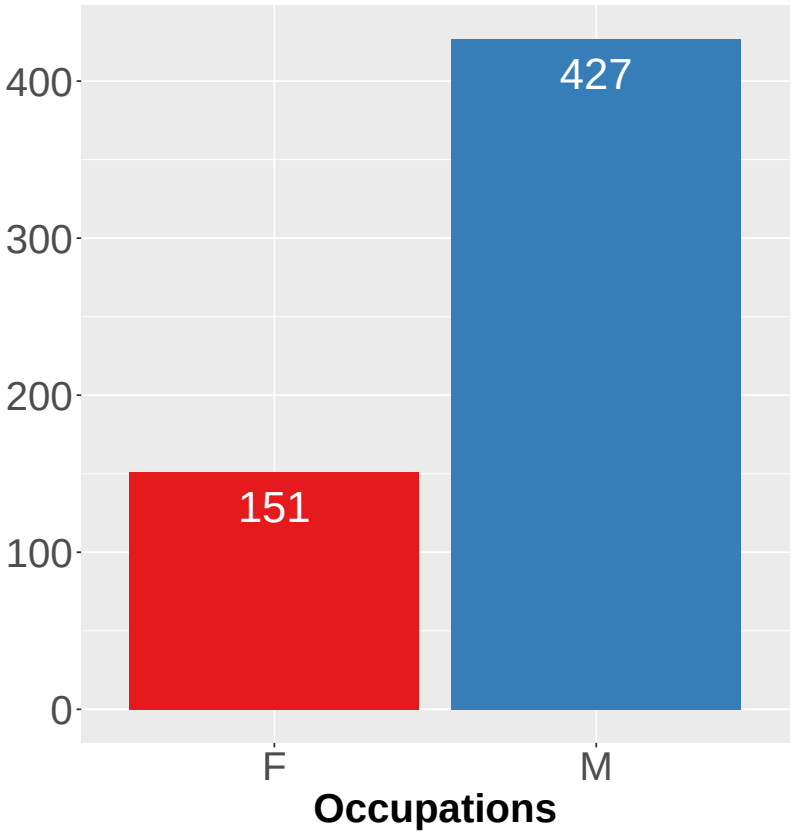


Top 5 Most Frequent Female Names

- 37% of all male names are ***John***.
  37% of all female names are ***Mary***.
- 3 of the top 5 male names are *John variants*: *John*, *Juan*, and *Ivan*.
  4 of the top 5 female names are "Mary" variants: *Mary, Maria, Marie*, and *Mari*.

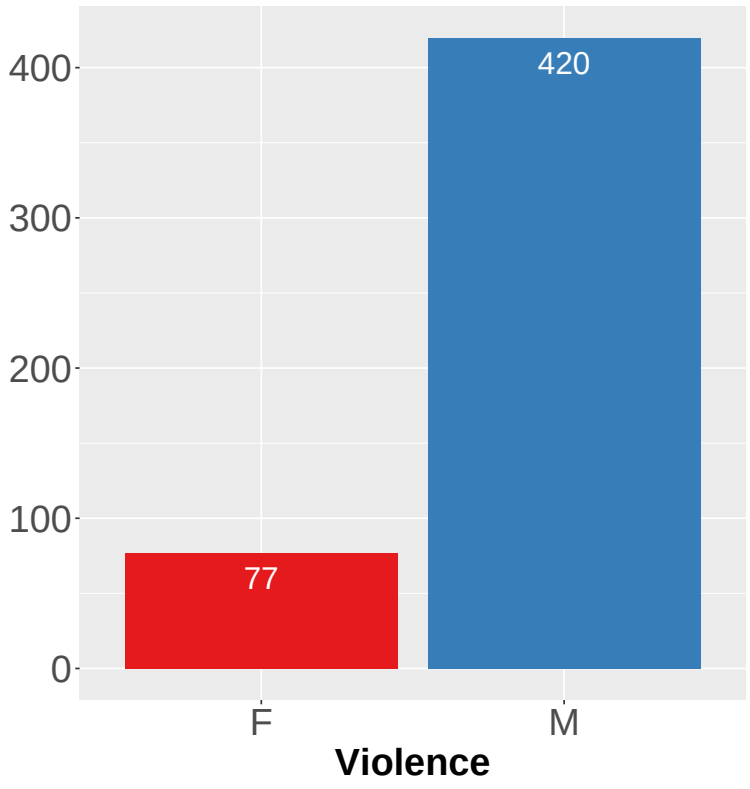
Occupations

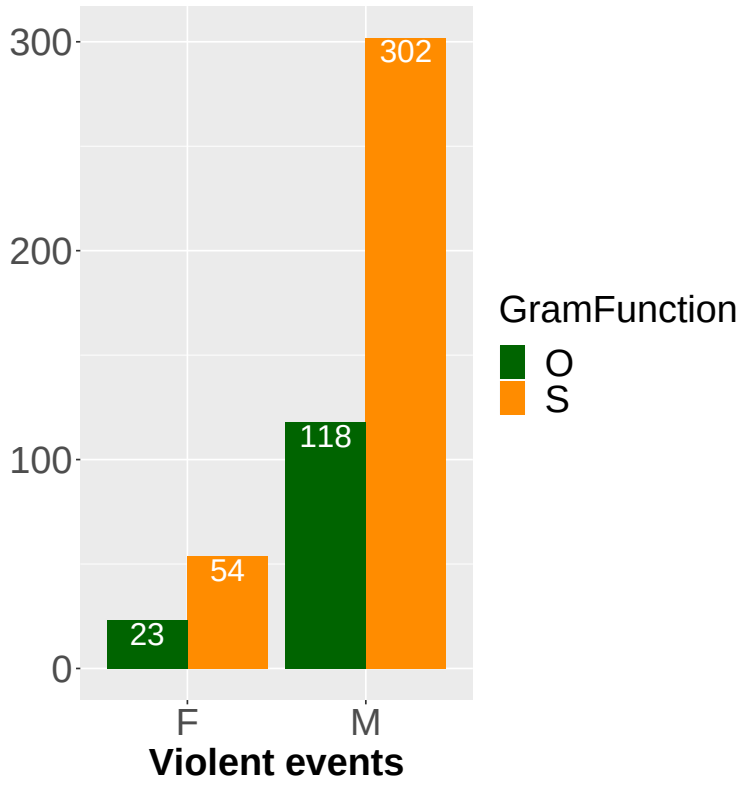
Male-gendered NPs are over-represented in occupation-related examples (74% M):



Violence

Male-gendered NPs are also massively over-represented in violence-related exx (85% M):

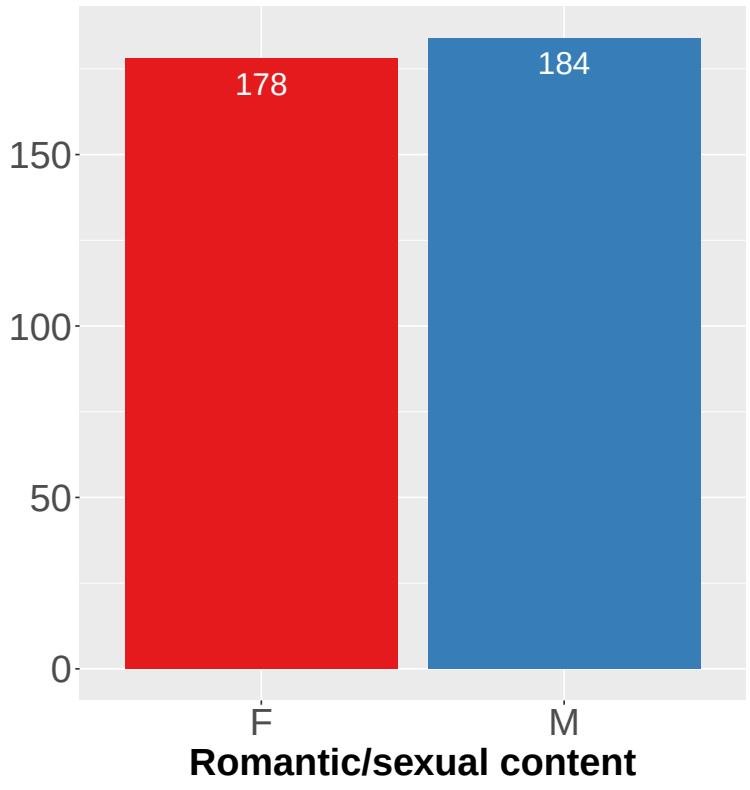


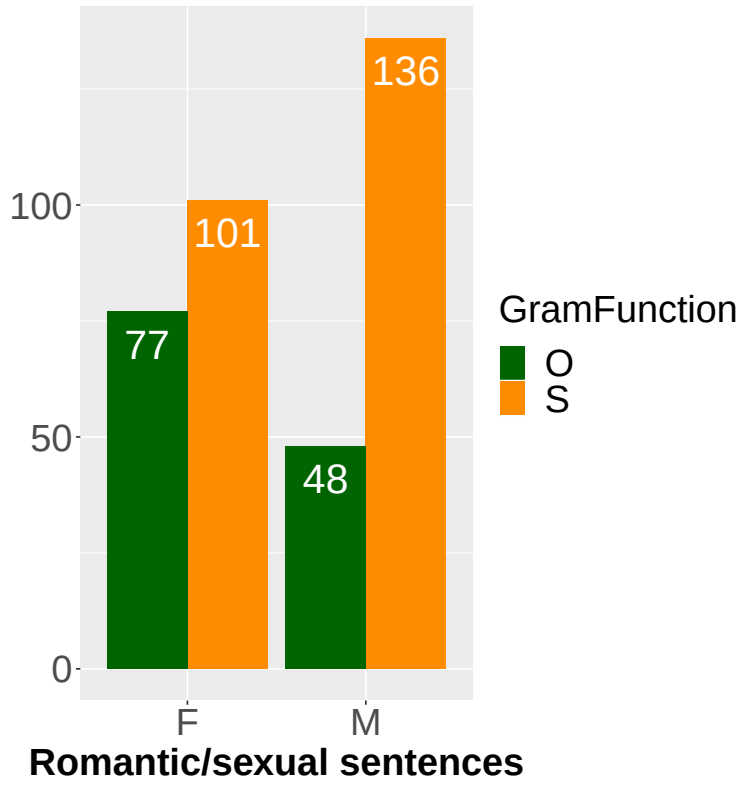


Female-gendered NPs: 70% subjects, Male-gendered NPs: 72% subjects.

Romance

Female-gendered NPs are over-represented in "romantic" exx (49% F):



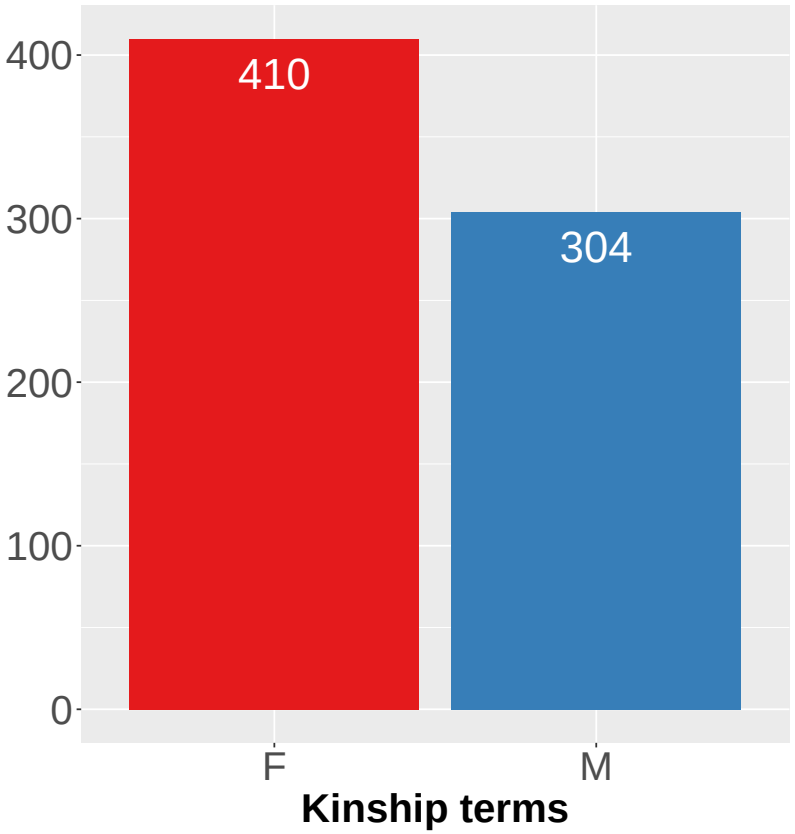


They are again over-represented as **objects** in such sentences:

Female-gendered NPs: 57% subjects, Male-gendered NPs: 73% subjects.

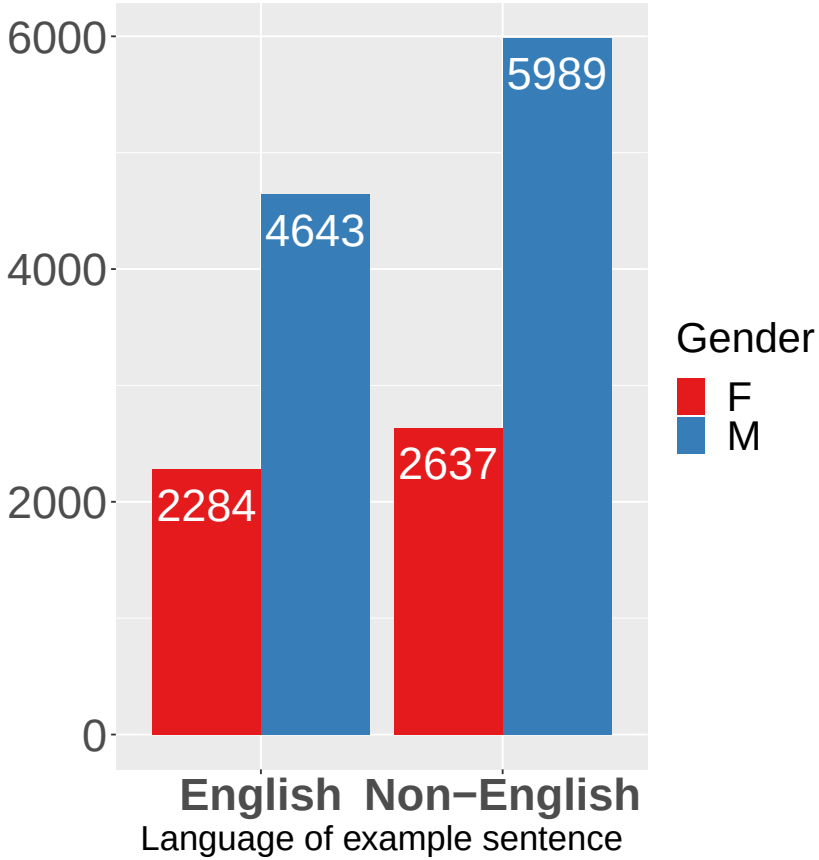
Kinship terms

Female-gendered NPs are massively over-represented with respect to kinship terms (57% F):



Language of examples

English and non-English examples don't appear different:



English: 33% female args

Non-English: 31% female args

Inappropriate examples

(3)
a. John ate the meal and Mary cleaned the dishes
b. John didn't eat the meal because he would have to clean the dishes
c. John thinks that he himself is a war hero
d. John told Bill that Mary began to cry without any reason
e. The boys had thrown no rock at the cars
f. Maria reviewed the novel, she didn't write it
g. The senator killed herself
h. Iraqi father drowned his 17 year old
i. Mary, being dumb, needs to sit down
j. Which Nobel prize winning author came in his car?

Inappropriate examples

(4)
a. For whom do you regret that she made a cake?
b. * Eat food that Mary$_i$ cooks, she$_i$ knows I never would
c. John (not Peter) washed cars well
d. * Kelly broke again tonight when she did the dishes
e. I called for a policeman, not a policewoman
f. The students are all the boys
g. Tomas replaced Ricardo as the captain
h. Every male student doesn't fear tigers
i. It is amazing how many cars he owns
j. At least one student of every professor$_i$ is horrified at his$_i$ grading procedure

Summary

Male-gendered NPs...

- appear more often as **arguments** than female-gendered NPs do
- appear more often as **subjects and agents** than female-gendered NPs do
- engage in significantly more **violence** than female-gendered NPs do
- have significantly more **occupations** than female-gendered NPs do

Summary

Female-gendered NPs...

- are over-represented as non-subjects, especially as recipients
- are over-represented in sentences involving romantic/sexual language
- are massively over-referred to using kinship terms

Summary

- Not (m)any suggestive or explicit examples
- …although stereotypes are very much evident (replicating Pabst et al. 2019)

$\rightsquigarrow$ We've tidied up the surface, but have done very little to address the underlying problem

- Language of example doesn't make a difference — so this effect is not (just) about lack of access to sources
- A slight improvement over the past 20 years: from low-30% to mid-30%

So what does this mean?

- Even if some ratios improved and explicit language has improved, **the gender bias observed in Macaulay and Brice and in Pabst et al. is still there**.
- Removing the blatant examples doesn't address the underlying problem.

We can do better!

Beyond the gender binary

- Explicit discussions of non-binary gender identities are **entirely absent**.
- There are also a number of other issues that should be addressed:
  - Western vs. Non-Western names
  - Constructed vs. corpus examples
  - Elicited examples, narratives, etc.

There's more to do

Remains to be done:

- Author gender
- Breakdown of non-English examples
- Sentiment analysis (positive/negative emotion)
- Predicates in:
  - violence
  - books and reading
  - appearances
  - romance-related
  - cars
  - other stereotypes

Your question here?

Roadmap

§1 Introduction

§2 Background: Macaulay and Brice (1997)

§3 Gender representation in journal papers

§4 **Discussion: Why does this matter? How can we improve?**

Why does this matter?

- **Constructed examples sentences are one of the main sources of data in theoretical linguistics**.
- **These examples are cited again and again**, often divorced from their original source and treated as an example from the literature of a particular phenomenon.
- **Examples may encode implicit biases** (even at a very subtle level), which then get handed down to new generations of linguists, perpetuating the cycle.

Why does this matter?

Inclusive language encourages participation from underrepresented groups

- leading to a **better community**
- leading to **better science**

...at the cost of just a little more thoughtfulness.

LSA Responses to Gender and Inclusivity

- 1996 Guidelines for Non-Sexist Usage (COSWL)
- 2016 Guidelines for Inclusive Language (COSWL)
- 2018 Panel at Annual Meeting: Our Linguistics Community: Addressing Bias, Power Dynamics, Harassment

What can we do?

In the interest of being maximally inclusive...

- Stereotypical language, sexually explicit and demeaning language, and language reflecting biases are easily avoidable, and should be.
- The use of gendered lexical items (-*man*, *he*, etc.) where unnecessary should be avoided.
- The biased and elevated frequency of particular gendered NPs in particular syntactic positions or semantic roles should be diminished.

What can we do?

Embrace singular *they*!

- We are often told that the pronoun *he* should be used for (singular) nouns whose gender is unknown.
- Despite this official designation, however, this pronoun feel exclusionary of non-male individuals.
- Singular *they* has been used for decades precisely for this purpose. #WOTY2019

Conclusion: What can we do?

- **Instructors:**
  - Choose your examples wisely.
  - Be sensitive to how you portray all individuals in your examples.
  - Keep in mind that you are in a position of authority and can have a *positive* influence on young minds entering the field.
  - Consider gender ratios and representation in your syllabi

- **Authors**
  - Be thorough, inclusive, and balanced in your citations.
  - Do not perpetuate bias in the examples you cite.
  - Keep the Guidelines for Inclusive Language in mind.

- **Editors/Reviewers**
  - Pay attention to the examples and language authors use.

Thank you!

Thank you! Questions?

Special thanks to **Monica Macaulay** and **Colleen Brice** for the original inspiration behind this work. Additional thanks to **Katarina Pabst**, **Kristen Syrett**, and **Paola Cèpeda**, whose work on the 2018 textbook project together with Hadas Kotek laid the ground for this project.

We are also thankful to the **Yale Female Faculty Forum** and **Claire Bowern** for supporting this work, and to our audience at Yale, NYU, MIT, and Brandeis University, as well as the 2018 LSA Annual Meeting, esp. Kirby Conrod, Emily Bender, and Lauren Ackerman, for their feedback on ideas which inspired this work.

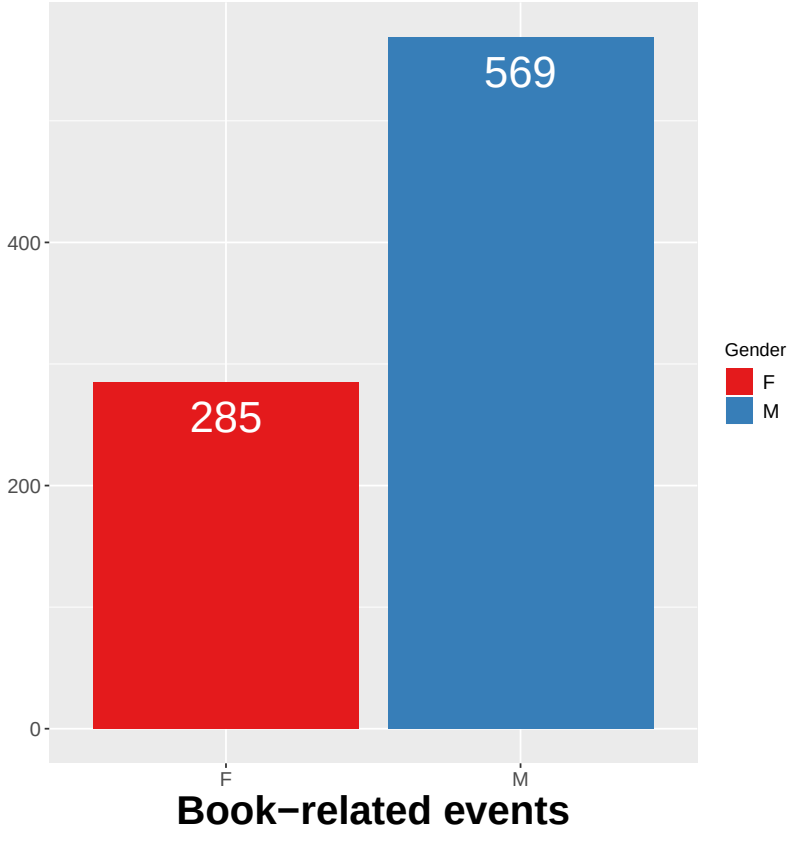
References I

Macaulay, Monica, and Colleen Brice. 1994. Gentlemen prefer blondes: A study of gender bias in example sentences. In *Cultural performances: Proceedings of the third Berkeley women and language conference*, 449–461.

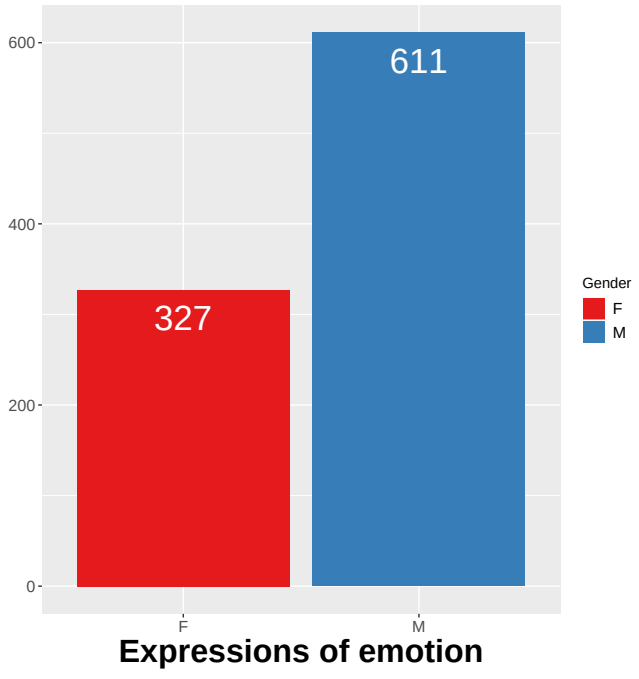
Macaulay, Monica, and Colleen Brice. 1997. Don't touch my projectile: gender bias and stereotyping in syntactic examples. *Language* 73:798–825.

Pabst, Katharina, Paola Cepeda, Hadas Kotek, and Kristen Syrett. 2019. Gender bias in linguistics textbooks: Has anything changed since Macaulay & Brice (1997)? Manuscript.

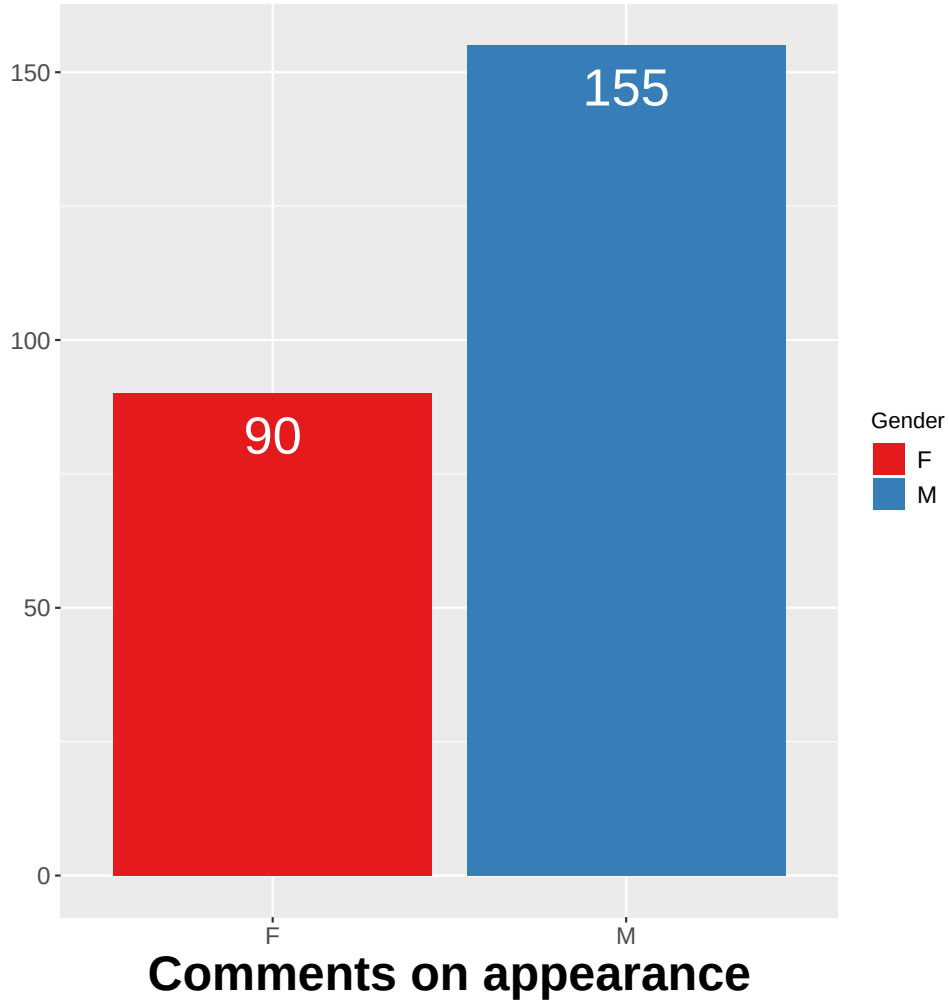
Gender in journal papers: Books and reading



Gender in journal papers: Emotion



Gender in journal papers: Appearance



the raw counts may not be the most helpful; the observation in M&B was about how often women were pleasing to men, so we need to know